Asymptotic Theory for Linear Mixed Effects Models With Large Cluster Size

ZIYANG LYU

A thesis submitted in fulfilment of
the requirements for the degree of
Doctor of Philosophy

Mathematical Sciences Institute
College of Physical and Mathematical Sciences
The Australian National University

June, 2020

Declaration

This thesis is an account of research undertaken between September 2015 and December 2019 at the Mathematical Sciences Institute, College of Physical and Mathematical Sciences, The Australian National University, Canberra, Australia.

Except where acknowledged in the customary manner, the material presented in this thesis is, to the best of my knowledge, original and has not been submitted in whole or part for a degree in any university.

Ziyang Lyu
June, 2020

Acknowledgements

I have been fortunate to have the unfailing support of a number of very talented and generous people over my time at the Australian National University.

Firstly, I would like to express my sincere gratitude to my supervisor, Professor Alan Welsh, for the continuous support of my Ph.D study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better supervisor and mentor for my Ph.D study.

Besides my supervisor, I would like to thank the rest of my advisory panel. Thanks especially to Associate Professor Robert Clark for his insightful comments and kindly assistance on simulations. Also, special thanks to Dr Francis Hui, whose occasional general assistance was invaluable over the last couple of years. From Mathematical Science Institute of the Australian National University, I am grateful to all administrative staff for assistance throughout my P.hd program, especially for Kelly Wicks and Mandy Gordon.

Last but not the least, I would like to thank my family: my parents and my wife for supporting me spiritually throughout writing this thesis and my life in general.

Abstract

This thesis first deals with asymptotic results for the maximum likelihood and restricted maximum likelihood (REML) estimators of the parameters in the nested error regression model when both the number of independent clusters and the cluster sizes (the number of observations in each cluster) go to infinity. A set of conditions is given under which the estimators are shown to be asymptotically normal. There are no restrictions on the rate at which the cluster size tends to infinity. Moreover, this thesis deals with the asymptotic distributions of the estimated best linear unbiased predictors (EBLUP) of the random effects, with ML/REML, estimated variance components, converge to the true distributions of the corresponding random effects, when both of the number of independent clusters and the cluster sizes (the number of observations in each cluster) go to infinity.

Contents

List of Tables

List of Figures

Chapter 1 Introduction

One of the goals of data analysts is to establish relationships between variables using regression models. The linear regression model is perhaps the simplest and most common model used in statistical analysis. Such a model summarizes the average relationship between a set of explanatory variables and the outcome/target variable. However, data rarely come from simple random samples. In social, behavioral, educational, and medical sciences, data is commonly collected via multistage clustered designs such as Raudenbush and Bryk (2002). The correlation between observations that belong to the same cluster should not be ignored as this will impact upon the computation of the standard errors of the regression parameters and hence on inference about these parameters. The extent of clustering is even more pronounced with longitudinal data. Longitudinal data are collected by cohort and panel surveys which have gained popularity. To appropriately take this type of data structure into account, the so-called linear mixed effects models are typically used. These models - as an extension of the linear regression model - contain additional random-effects.

1.1 Linear Mixed Models

Suppose that from a set of g independent clusters, we observe clustered pairs $(\mathbf{y}_i, \mathbf{X}_i)$, where $\mathbf{y}_i$ represents a $m_i \times 1$ vector of continuous responses for the i-th cluster. We present elements of the $\mathbf{y}_i$ vector as follows,

$$\mathbf{y}_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{im_i} \end{bmatrix}.$$

Note that the number of elements, m_i in the vector $\mathbf{y}_i$ may vary from one cluster to another. The number of observations is then given by $n = \sum_{i=1}^{g} m_i$.

Turning to the covariate, let $\mathbf{X}_i$ denote an $m_i \times p$ design matrix representing the known values of the p covariates for the ith cluster. We write

$$\mathbf{X}_i = [\mathbf{X}_i^{(1)}, \mathbf{X}_i^{(2)}, \ldots, \mathbf{X}_i^{(p)}] = \begin{bmatrix} X_{1i}^{(1)} & X_{1i}^{(2)} & \ldots & X_{1i}^{(p)} \\ X_{2i}^{(1)} & X_{2i}^{(2)} & \ldots & X_{2i}^{(p)} \\ \vdots & \vdots & \vdots & \vdots \\ X_{m_i i}^{(1)} & X_{m_i i}^{(2)} & \ldots & X_{m_i i}^{(p)} \end{bmatrix},$$

where $\mathbf{X}_i^{(p)}$ is the p-th column of matrix $\mathbf{X}_i$. In a model including an intercept term, the first column would be set equal to 1 for all observations. For ease of presentation, we assume that the $\mathbf{X}_i$ matrices are of full rank; that is, none of the columns (or rows) is a linear combination of the others. In general, $\mathbf{X}_i$ matrices may not be of full rank, and this may lead to an aliasing (or parameter identifiability) problem for the fixed effects stored in the vector $\boldsymbol{\beta}$.

The linear mixed model is a statistical model containing both fixed effects and random effects. With the above observed data, a linear mixed model may be expressed as

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\alpha} + \mathbf{e}_i, \qquad i = 1, \ldots, g. \tag{1.1.1}$$

The $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ in (1.1.1) is a $p \times 1$ vector of fixed effects associated with the p covariates used in constructing the $\mathbf{X}_i$ matrix. The $m_i \times q$ matrix $\mathbf{Z}_i$ in (1.1.1) is a design matrix that represents the known values of the q covariates. This matrix is very much like the $\mathbf{X}_i$ in that it contains the observed values of covariates; however, it usually has fewer columns than the $\mathbf{X}_i$ matrix i.e., $q < p$,

$$\mathbf{Z}_i = [\mathbf{Z}_i^{(1)}, \mathbf{Z}_i^{(2)}, \ldots, \mathbf{Z}_i^{(q)}] = \begin{bmatrix} Z_{1(i)}^{(1)} & Z_{1(i)}^{(2)} & \ldots & Z_{1(i)}^{(q)} \\ Z_{2(i)}^{(1)} & Z_{2(i)}^{(2)} & \ldots & Z_{2(i)}^{(q)} \\ \vdots & \vdots & \vdots & \vdots \\ Z_{m_i(i)}^{(1)} & Z_{m_i(i)}^{(2)} & \ldots & Z_{m_i(i)}^{(q)} \end{bmatrix},$$

where $\mathbf{Z}_i^{(q)}$ is the q-th column of matrix $\mathbf{Z}_i$. The columns in the $\mathbf{Z}_i$ matrix represent observed values for the q predictor variables for the i-th cluster, which have effects on the continuous response variable that vary randomly across clusters i.e., cluster specific effects. In many cases, predictors with effects that vary randomly across clusters are represented in both the $\mathbf{X}_i$ matrix and the $\mathbf{Z}_i$ matrix. In a linear mixed model in which only the intercepts are assumed to vary randomly across subjects, the $\mathbf{Z}_i$ is a column of 1's.

The $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_q)^T$ is a $q \times 1$ vector of random effects associated with the q covariates in the $\mathbf{Z}_i$ matrix. Recall that by definition, random effects are random variables. In many studies, it is assumed that the $\boldsymbol{\alpha}$ vector follows a multivariate normal distribution, with mean vector $\mathbf{0}$ and a variance-covariance matrix denoted by $\mathbf{D}$. However, in practice, we can never be sure that the assumption of normality holds. In this thesis, we consider more general situations where $\boldsymbol{\alpha}$ is not necessarily required to be normal distributed.

Finally, the $\mathbf{e}_i = (e_{i1}, \ldots, e_{im_i})^T$ in (1.1.1) is an m_i vector of residuals. The element e_{ij} represents the residual associated with an observed response at the j-th occasion for the i-th cluster. Note that the $\mathbf{e}_i$ vectors may have a different number of elements. In contrast to the standard linear model, the residuals associated with repeated observations on the same subject in a linear mixed model can be correlated. Specifically, we assume the vector of residuals $\mathbf{e}_i$ follows a multivariate distribution with mean vector $\mathbf{0}$ and a positive-definite symmetric variance-covariance matrix $\mathbf{R}_i$. In many cases, we assume $\mathbf{R}_i$ is a diagonal matrix, under normality, which means the residuals within a cluster are independent. We usually parameterize $\mathbf{D}$ and $\mathbf{R}_i$ by $\boldsymbol{\theta}$, a $c \times 1$ vector of parameters, in the same way. In this case, we write $\mathbf{D}(\boldsymbol{\theta})$ and $\mathbf{R}_i(\boldsymbol{\theta})$.

We denote $\mathbf{y}^T = [\mathbf{y}_1^T, \ldots, \mathbf{y}_g^T]$, and then the linear mixed model is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \mathbf{e}, \tag{1.1.2}$$

where

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_g \end{bmatrix}, \qquad \mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 \\ \vdots \\ \mathbf{Z}_g \end{bmatrix}, \qquad \mathbf{e} = \begin{bmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_g \end{bmatrix}, \qquad \text{and} \qquad \mathbf{R} = \mathrm{diag}[\mathbf{R}_1 \ldots, \mathbf{R}_g].$$

The linear mixed model is a powerful modeling tool for statistical analysis, which has gathered increasing popularity over the past two decades. The fixed effects $\boldsymbol{\beta}$ may be associated with continuous covariates such as weight, baseline test score, or socioeconomic status, which take on values from a continuous range, or with factors, such as gender or treatment group, which are categorical. Fixed effects are unknown constant parameters associated with continuous covariates or the levels of categorical factors in a linear mixed model. The estimation of these population averaged parameters in linear mixed models is generally of intrinsic interest, because they indicate the relationships of the covariates with the continuous outcome variable. In regression models, fixed effects are known as regression coefficients.When the levels of a categorical factor can be thought of as having randomly samples (e.g., classrooms or clinics that are randomly sampled from a larger population of classes or clinics), the effects associated with levels of those factors can be modeled as random effects in a linear mixed model. In contrast to fixed effects, random effects help to reduce the number of parameters in the model. Moreover, random effects induce a correlation between y_{ij} and y_{ik} for $j \neq k$ in linear mixed models with random effects. But, y_{ij} and y_{ik} are independent of each other in a fixed-effect model. Finally, linear mixed models often produce better predictions than purely fixed effect models.

1.2 Balanced Data and Unbalanced Data

1.2.1 Balanced Data

Data can often be usefully characterized in several ways that depend on whether or not each cluster contains the same number of observations. When these counts are the same, the data are described as balanced data; they typically come from designed factorial experiments that have been executed as planned. An explicit definition of a very broad class of balanced data is given in Searle (1987). The analysis of balanced data, whether the models include random effects or not, and whether there are interactions or not, is comparably straightforward and has been extensively studied. In chapter 2, we start from balanced data (each cluster size is same), which is much easier for obtaining desirable asymptotic property results. This is because balanced data typically allows the use of Kronecker products, which simplifies derivations in many formulas and equations.

1.2.2 Unbalanced Data

After defining balanced data, we are left with unbalanced data. This is data where the numbers of observations in clusters are not all equal, and may in fact be quite unequal. Survey data are often like this, where data are often collected outside of a controlled experimental setting because they exist and so the numbers of observations in the clusters are simply those that are available. Records of many human activities are of this nature; e.g., yearly income for people classified by age, sex, education, education of each parent, and so on. Tippett (1952) indicated that in summing the squares of the deviations of the cluster means from the grand mean, each cluster has been given a different weight m_i (the number of observations in the i-th cluster). Unbalanced data are usually more complex to derive large sample properties. Nevertheless, in chapter 3, we extend to considering the general case (unbalanced data) to study the asymptotic properties of Maximum likelihood (ML) and restricted maximum likelihood (REML) in linear mixed models. Such results are more complicated to derive because we cannot use the elegant structure from balanced data to obtain the relevant results.

1.3 Estimation in LMM

In the theory of linear mixed effects models, when interest focuses on the accurate estimation of the variance components (Searle et al., 1992), restricted/residual maximum likelihood (REML) theory might be the preferred option (Verbeke and Molenberghs, 2000). For a valid interpretation of the parameters of linear mixed effects models (regression coefficients and variance components) asymptotic properties/ results of these estimators are important. In a linear mixed model, we estimate the fixed-effect parameters, $\boldsymbol{\beta}$, and the covariance parameters $\boldsymbol{\theta}$. In this section, we discuss maximum likelihood (ML) and restricted maximum likelihood (REML) estimation, which are methods commonly used to estimate these parameters.

1.3.1 Maximum Likelihood (ML) Estimation

In general, maximum likelihood (ML) estimation is a method of obtaining estimates of unknown parameters by maximizing a likelihood function. To apply ML estimation, we make distributional assumptions that allow us to construct the likelihood as a function of the parameters in the specified model. Maximum likelihood estimates (MLEs) of the parameters are the values of the arguments that maximize the likelihood function (i.e., the values of the parameters that make the observed values of the dependent variable most likely, given the distributional assumptions). We refer to Casella and Berger (2002) for an in-depth discussion of ML estimation.

In the context of the linear mixed model, we construct the likelihood function of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ under the Gaussian marginal model $\mathbf{y}_i \sim N(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{V}_i(\boldsymbol{\theta}))$ in (1.1.1), where $\mathbf{V}_i(\boldsymbol{\theta}) = \mathbf{Z}_i\mathbf{D}(\boldsymbol{\theta})\mathbf{Z}_i^T + \mathbf{R}_i(\boldsymbol{\theta})$. Note that $\mathbf{V}_i(\boldsymbol{\theta})$, $\mathbf{D}(\boldsymbol{\theta})$ and $\mathbf{R}_i(\boldsymbol{\theta})$ are functions of the covariance parameters in $\boldsymbol{\theta}$. Given the observed data, the likelihood function contribution for the ith cluster is defined as

$$L_i(\boldsymbol{\beta}, \boldsymbol{\theta}; \mathbf{y}_i) = (2\pi)^{-\frac{m_i}{2}} |\mathbf{V}_i(\boldsymbol{\theta})|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^T \mathbf{V}_i(\boldsymbol{\theta})^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})\right\}, \tag{1.3.1}$$

where $|\mathbf{V}_i(\boldsymbol{\theta})| = \det[\mathbf{V}_i(\boldsymbol{\theta})]$ is the determinant of $\mathbf{V}_i(\boldsymbol{\theta})$. The likelihood function for all the clusters, denoted here as $L(\boldsymbol{\beta}, \boldsymbol{\theta})$. Note that $L(\boldsymbol{\beta}, \boldsymbol{\theta})$ is the product of the g independent contributions defined in (1.3.1) for g individuals $(i = 1, \ldots, g)$:

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}) = \prod_{i=1}^{g} L_i(\boldsymbol{\beta}, \boldsymbol{\theta}) = \prod_{i=1}^{g} (2\pi)^{-\frac{m_i}{2}} |\mathbf{V}_i(\boldsymbol{\theta})|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^T \mathbf{V}_i^{-1}(\boldsymbol{\theta})(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})\right\}. \tag{1.3.2}$$

The corresponding log-likelihood function, $l(\boldsymbol{\beta}, \boldsymbol{\theta})$ is

$$l(\boldsymbol{\beta},\boldsymbol{\theta}) = \log L(\boldsymbol{\beta},\boldsymbol{\theta}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{g}\log|\mathbf{V}_i(\boldsymbol{\theta})| - \frac{1}{2}\sum_{i=1}^{g}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^T\mathbf{V}_i^{-1}(\boldsymbol{\theta})(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}), \tag{1.3.3}$$

where $n = \sum_{i=1}^{g} m_i$ is the number of observations. We can estimate $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ simultaneously, by optimization of $l(\boldsymbol{\beta}, \boldsymbol{\theta})$ with respect to both $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. It can be shown that the estimating equation is equivalent to

$$\begin{aligned} \boldsymbol{\beta} &= \left\{\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}\right\}^{-1}\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{y} \\ (\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T\mathbf{V}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_k}\mathbf{V}^{-1}(\boldsymbol{\theta})(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) &= \operatorname{tr}\left\{\mathbf{V}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_k}\right\}, \quad k = 1,\ldots,c. \end{aligned}$$

Provided that $\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}$ is of full rank, we can write $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \{\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}\}^{-1}\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{y}$. Given a specific parametric form for $\mathbf{D}(\boldsymbol{\theta})$ and $\mathbf{R}(\boldsymbol{\theta})$, and hence for $\mathbf{V}(\boldsymbol{\theta})$, then we can specify the partial derivatives of $\frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_k}$, where $\boldsymbol{\theta}_k$ is the k-th element of vector $\boldsymbol{\theta}$.

1.3.2 Restricted Maximum likelihood (REML) Estimation

REML estimation is an alternative way of estimating the covariance parameters in $\boldsymbol{\theta}$ and then the fixed effect parameters $\boldsymbol{\beta}$s. REML estimation (sometimes called residual maximum likelihood estimation) was first introduced in the early 1970s by Patterson and Thompson (1971) as a method of estimating variance components in the context of unbalanced incomplete block designs. Alternative and more general derivations of REML are given by Harville (1977), and Verbyla (1990).

In linear mixed models, REML estimation is often preferred to ML estimation, because it produces less biased estimates of covariance parameters by taking into account the loss of degrees of freedom that results from estimating the fixed effects in $\boldsymbol{\beta}$. The REML estimates of $\boldsymbol{\theta}$ are obtained by maximizing the REML criterion function

$$l_R(\boldsymbol{\theta}) = -\frac{n-p}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{g}\log|\mathbf{V}_i(\boldsymbol{\theta})| - \frac{1}{2}\sum_{i=1}^{g}\mathbf{r}_i^T(\boldsymbol{\theta})\mathbf{V}_i^{-1}(\boldsymbol{\theta})\mathbf{r}_i(\boldsymbol{\theta}) - \frac{1}{2}\sum_{i=1}^{g}\log|\mathbf{X}_i^T\mathbf{V}_i^{-1}(\boldsymbol{\theta})\mathbf{X}_i|, \tag{1.3.4}$$

where $\mathbf{r}_i(\boldsymbol{\theta}) = \mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ with $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \left\{\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}\right\}^{-1}\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{y}$. The REML criterion function is a modified profile log-likelihood function for $\boldsymbol{\theta}$. We estimate $\boldsymbol{\theta}$ by maximizing $l_R(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ and then estimate $\boldsymbol{\beta}$ by substituting the estimate of $\boldsymbol{\theta}$ into $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$. It can be shown that REML equation,

$$\mathbf{y}^T\mathbf{P}(\boldsymbol{\theta})\frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_k}\mathbf{P}(\boldsymbol{\theta})\mathbf{y} = \operatorname{tr}\left\{\mathbf{P}(\boldsymbol{\theta})\frac{\partial \mathbf{V}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_k}\right\}, \quad k = 1,\ldots,c,$$

where

$$\mathbf{P}(\boldsymbol{\theta}) = \mathbf{V}^{-1}(\boldsymbol{\theta}) - \mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}\left\{\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}\right\}^{-1}\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta}).$$

Note that the REML estimator of $\boldsymbol{\beta}$ is of the same form as the MLE of $\boldsymbol{\beta}$ except that $\boldsymbol{\theta}$ is replaced by its REML estimator.

1.4 Prediction in LMM

Prediction in linear mixed models has a long history, starting with Henderson (1953) in his early work in the field of animal breeding. Arguably the best known approach to prediction of random effects in linear mixed models is given by the best linear unbiased prediction (BLUP) method. Robinson (1991) gives a wide-ranging account of the BLUP with examples and applications.

When $\boldsymbol{\theta}$ is known, BLUPs of the realized values of the random variables $\boldsymbol{\alpha}_i$ are linear in the sense that they are linear functions of the data, $\mathbf{y}$; unbiased in the sense that the expected value of the estimate is equal to the expected value of the quantity being estimated and best in sense that they have minimum mean squared error within the class of linear unbiased estimators. They are also predictors in order to distinguish them from estimators of fixed effects. A convention has developed that estimators of random effects are called predictors while estimators of fixed effects are called estimators. Mathematically, the BLUPs $\tilde{\boldsymbol{\alpha}}_i$ of $\boldsymbol{\alpha}_i$ were defined as solutions to the mixed model equations by Henderson (1975). We write BLUPs,

$$\tilde{\boldsymbol{\alpha}}(\boldsymbol{\theta}) = \mathbf{D}(\boldsymbol{\theta})\mathbf{Z}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\{\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})\},$$

where $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \{\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}\}^{-1}\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{y}$.

Of course, the variance components are typically unknown. In linear mixed models, it is customary to replace the vector, $\boldsymbol{\theta}$, which is involved in the expression of BLUP by a consistent estimator, $\hat{\boldsymbol{\theta}}$. The resulting predictor is often called an estimated BLUP, or EBLUP. In this thesis, we focus on the EBLUPs referring to random effects. Kackar and Harville (1981) showed that, if $\hat{\boldsymbol{\theta}}$ is an even (e.g., $h(u) = h(-u)$) and translation invariant (e.g., $h(u + a) = h(u)$) estimator and the data are normal, the EBLUP remains unbiased. Some of the well-known estimators of $\boldsymbol{\theta}$, including Analysis of variance (ANOVA), ML (discussed in section 1.3.1) and REML (discussed in section 1.3.2) estimators, are even and translation invariant . In their arguments, however, Kackar and Harville assumed the existence of the expected value of EBLUP, which is not immediately evident because unlike BLUP, EBLUP is no longer linear in $\mathbf{y}$. The existence of the expected value of EBLUP was only proven later on by Jiang (1999). Harville (1977) considered the one-way random-effects model, and showed that the EBLUP of the mixed effect equals a parametric empirical Bayes (PEB) estimator for the normal model.

Measuring the uncertainty of EBLUP gives a sentence as to why it is an important problem. A standard measure is a prediction mean squared error, such as $\mathrm{E}\{\tilde{\boldsymbol{\alpha}}(\boldsymbol{\theta}) - \boldsymbol{\alpha}\}\{\tilde{\boldsymbol{\alpha}}(\boldsymbol{\theta}) - \boldsymbol{\alpha}\}^T$. However, unlike the BLUP, the prediction mean squared error of the EBLUP usually does not have an analytic expression. The mean squared error of the EBLUP may be obtained by first deriving the mean squared error of the BLUP and then replacing the unknown $\boldsymbol{\theta}$ in the expression by $\hat{\boldsymbol{\theta}}$. However, this is an under-estimate of the true mean squared error of the EBLUP. Because Kackar and Harville (1984), Prasad and Rao (1990) and Jiang et al. (2002) study serval methods of approximation to the mean squared error of the EBLUP.

Harville noted that much of the work on parametric empirical Bayes has been of a theoretical nature. The work has tended to focus on relatively simple models, such as the one-way random effects model because these simple models are tractable from a theoretical standpoint. On the other hand, much of the work on EBLUP has been carried out by practitioners such as researchers in the animal breeding area and has been applied to relatively complex areas.

1.5 Example

A good way to understand linear mixed models is considering simple special cases. Consider a nested error regression model with a single covariate. In this example, the linear mixed model can be written as

$$y_{ij} = \mu(x_{ij}) + \alpha_i + e_{ij} \qquad \text{for} \qquad i = 1, \ldots, g; \quad j = 1, \ldots, m_i. \tag{1.5.1}$$

In this model, y_{ij} is the j-th observation in the i-th cluster, and $\mu(x_{ij}) = \beta_0 + x_{ij}\beta_1$. The quantity x_{ij} is a known covariate, and β_0, β_1 are unknown fixed effects parameters. The quantity e_{ij} is an error with mean 0 and variance σ_e^2, while α_i is a random variable with mean 0 and variance σ_α^2. Relating to (1.1.1), let $q = 1$, $p = 2$, $\mathbf{X}_i = [\mathbf{1}_{m_i}, \mathbf{x}_i]$,where $\mathbf{x}_i = (x_{i1}, \ldots, x_{im_i})^T$, and $\mathbf{Z}_i$ be a $m_i \times q$ matrix with all zero elements but the i-th column being a column of 1's. So we have $\mathbf{D} = \mathbf{I}_g\sigma_\alpha^2$ and $\mathbf{R}_i = \mathbf{I}_{m_i}\sigma_e^2$ in (1.5.1). This model reduces to the usual regression model when $\sigma_\alpha^2 = 0$ or the α_i are fixed unknown parameters. We use the following example to illustrate these differences between the linear regression model and the linear mixed model.

Example 1. Suppose a new form of injectable insulin is being tested using g patients. It is reasonable to think of those patients as a randomly chosen sample of patients from a population of patients in a city. If patient i is tested m_i times in the trial ($n = \sum_{i=1}^{g} m_i$) and the measured response of j-th tests is denoted by y_{ij}, then it is not unreasonable to assume that

$$\mathrm{E}(y_{ij}) = \mu_i,$$

where μ_i is the expected blood-sugar from the observed patient i. We write $\mu_i = \mu + \alpha_i$, where μ is a general mean and α_i is the effect on blood-sugar level of the observed patient i, and consider the cases α_i are fixed unknown parameters and α_i are independent random variables.

In terms of the linear regression model, the expected value of y_{ij}, i.e., μ_i (or μ and each α_i) is considered as a fixed unknown constant. We can then write $\mu_i = \mu + \alpha_i$ as fixed effect, we can write

$$y_{ij} = \mu + \alpha_i + e_{ij} \qquad \text{for} \qquad i = 1, \ldots, g; \quad j = 1, \ldots, m_i, \tag{1.5.2}$$

where α_i is fixed, satisfying $\sum_{i=1}^{g} \alpha_i = 0$ for parameter identifiability. Thus, we have

$$\mathrm{Cov}(y_{ij}, y_{i'j'}) = \begin{cases} 0 & \text{if } i \neq i' \\ 0 & \text{if } i = i', j \neq j' \\ \sigma_e^2 & \text{if } i = i', j = j' \end{cases},$$

which means $\{y_{ij}\}$ are independent with each other. When we estimate α_i in the linear regression model, we have

$$\hat{\alpha}_i = \bar{y}_{i.} - \bar{y}_{..} = \alpha_i - \bar{\alpha} + \bar{e}_{i.} - \bar{e}_{..}.$$

where $\bar{y}_{i.} = \frac{1}{m_i}\sum_{j=1}^{m_i} y_{ij}$, $\bar{y}_{..} = \frac{1}{n}\sum_{i=1}^{g}\sum_{j=1}^{m_i} y_{ij}$, $\bar{\alpha} = \frac{1}{g}\sum_{i=1}^{g}\alpha_i$, $\bar{e}_{i.} = \frac{1}{m_i}\sum_{j=1}^{m_i} e_{ij}$ and $\bar{e}_{..} = \frac{1}{n}\sum_{i=1}^{g}\sum_{j=1}^{m_i} e_{ij}$. Furthermore, we have

$$\mathrm{Var}(\hat{\alpha}_i) = \mathrm{Var}(\alpha_i - \bar{\alpha} + \bar{e}_{i.} - \bar{e}_{..}) = \mathrm{Var}(\bar{e}_{i.} - \bar{e}_{..}) = (\frac{1}{m_i} - \frac{1}{n})\sigma_e^2. \tag{1.5.3}$$

The assumption of independence between the y_{ij}'s, which underlies the linear regression model, may not be a reasonable one. This is because there exists correlations among the observations from the same individual, especially if the times at which observations are collected are relatively close. Moreover, the linear regression model allow us to make inference on the levels occurring in the data that we obtain. However, our patients were taken as a random sample of patients from a population of patients in a city, and we are interested in the whole population and not just these patients inferences we seek. Thus it is not appropriate to apply inference based on linear regression model to the population.

In terms of the linear mixed model, we take α_i as a random effect of blood-sugar level related to observed patient i, who is one from among a population of possible patients. Since these patients have been chosen randomly with the objective being to treat them as representative of the population of all patients in the city, then the one labeled i is of no particular interest in itself to the trial. This then motivates the linear mixed model

$$y_{ij} = \mu + \alpha_i + e_{ij} \qquad \text{for} \qquad i = 1, \ldots, g; \quad j = 1, \ldots, m_i, \tag{1.5.4}$$

where the α_i's are independently and identically distributed (i.i.d.) with mean 0 and variance σ_α^2. Although (1.5.4) is algebraically the same as (1.5.2), the assumptions underlying it are different. Thus, we have

$$\text{Cov}(y_{ij}, y_{i'j'}) = \begin{cases} 0 & \text{if } i \neq i' \\ \sigma_\alpha^2 & \text{if } i = i', j \neq j' \\ \sigma_\alpha^2 + \sigma_e^2 & \text{if } i = i', j = j' \end{cases},$$

We see that patients are independent, but there are correlation in observations collected from the same patient. Furthermore, we can predict α_i in the linear mixed model using the EBLUP as

$$\hat{\alpha}_i = \hat{\eta}_i(\bar{y}_{i.} - \bar{y}_{..}),$$

where $0 \leq \hat{\eta}_i = \frac{m_i\hat{\sigma}_\alpha^2}{\hat{\sigma}_e^2 + m_i\hat{\sigma}_\alpha^2} \leq 1$ and $\hat{\sigma}_\alpha^2$ and $\hat{\sigma}_e^2$ are estimators of σ_α^2 and σ_e^2. The prediction mean squared error for $\hat{\alpha}_i$ is much more complicated than (1.5.3).

1.6 Variance of limit or limit of variance

When Kackar and Harville (1984), Prasad and Rao (1990) and Jiang et al. (2002) approximated asymptotic MSE of EBLUPs, they used the limit of the MSE, which leads to very complicated calculations and restrictive conditions. They were partly forced to take this approach because they did not make use of $m_L \to \infty$ in their analysis, where m_L is lower bound of m_i, the i-th cluster size. The results in Chapter 2-3 allow us to take advantage of the simplification when $m_L \to \infty$ with $g \to \infty$. Thus, we try to explore another approach , taking the MSE of a linear approximation rather than starting from approximating the asymptotic MSE.

We consider a simple example to illustrate the two different ways of approximating asymptotic variances. This is a special case of estimating the asymptotic MSE when asymptotic bias is zero but gives useful insight into the general case. Suppose $g(\hat{\theta})$ is a function of the estimator $\hat{\theta}$ that estimates θ_0. Suppose $\hat{\theta} - \theta_0$ satisfies the central limit theorem so $n^{\frac{1}{2}}(\hat{\theta} - \theta_0) \xrightarrow{D} N(0, \sigma^2)$. Note that $\hat{\theta}$, θ_0 n and σ here are not the same as those used in previous sections, and they are more generic notations being used here. We want to estimated $n\,\text{Var}\left\{g(\hat{\boldsymbol{\theta}})\right\}$. If this quantity is complicated, some researchers take the limit of the variance, $\lim_{n\to\infty} n\,\mathrm{Var}\{g(\hat{\theta})\}$, and then estimate the limiting form. This involves first calculating the variance (moment) of $g(\hat{\theta})$, where $\hat{\theta}$ is an estimator of θ_0, and then finding the limit of the variance. However, this method often requires complicated calculations. Thus instead of the limit of the variance, we take the variance of a limiting linear approximation. In this section, we discuss the difference between these two approaches.

1.6.1 Limit of variance

We have an exact expression for $g(\hat{\theta})$,

$$g(\hat{\theta}) = g(\theta_0) + g'(\theta_0)(\hat{\theta} - \theta_0) + R(\hat{\theta}), \tag{1.6.1}$$

where $R(\hat{\theta})$ is the remainder. Next, we calculate the variance of $g(\hat{\theta})$ as

$$\begin{aligned}
\mathrm{Var}\{g(\hat{\theta})\} &= \mathrm{Var}\{g(\theta_0) + g'(\theta_0)(\hat{\theta} - \theta_0) + R(\hat{\theta})\} \\
&= \mathrm{Var}\{g'(\theta_0)(\hat{\theta} - \theta_0) + R(\hat{\theta})\} \\
&= \mathrm{Var}\{g'(\theta_0)(\hat{\theta} - \theta_0)\} + \mathrm{Var}\{R(\hat{\theta})\} + 2\,\mathrm{Cov}\{g'(\hat{\theta}_0)(\theta - \theta_0), R(\hat{\theta})\}
\end{aligned}$$

If we impose some conditions here to simplify $\mathrm{Var}\{g(\hat{\theta})\}$ by ensuring that $\mathrm{Cov}\{g'(\theta_0)(\hat{\theta} - \theta_0), R(\hat{\theta})\} = 0$, then we can take the limit as

$$\begin{aligned}
\lim_{n\to\infty} n\,\mathrm{Var}\{g(\hat{\theta})\} &= \lim_{n\to\infty} n\,\mathrm{Var}\{g'(\theta_0)(\hat{\theta} - \theta_0)\} + \lim_{n\to\infty} n\,\mathrm{Var}\{R(\hat{\theta})\} \\
&= g'(\theta_0)^2\sigma^2 + \lim_{n\to\infty} n\,\mathrm{Var}\{R(\hat{\theta})\}.
\end{aligned}$$

If we set additional conditions to ensure that $\lim_{n\to\infty} n\,\mathrm{Var}\{R(\hat{\theta})\} = o(1)$, we can obtain

$$n\,\mathrm{Var}\left\{g(\hat{\theta})\right\} \to g'(\theta_0)^2\sigma^2.$$

1.6.2 Variance of linear approximation

Turning to the variance of linear approximation, we first need to set some conditions to ensure that the remainder term in a limit of function $g(\hat{\theta})$ converges to zero in probability. We write

$$\{g(\hat{\theta}) - g(\theta_0)\} = g'(\theta_0)(\hat{\theta} - \theta_0)\{1 + o_p(1)\}.$$

By Slutsky's theorem, we obtain

$$n^{\frac{1}{2}}\{g(\theta) - g(\theta_0)\} \xrightarrow{D} N(0, g'(\theta_0)^2\sigma^2)$$

and then we use the asymptotic variance $g'(\hat{\theta})^2\sigma^2$ to approximate the variance of $n^{\frac{1}{2}}g(\hat{\theta})$.

We find that the biggest difference between the two approaches is in which step we set conditions to simplify the approximation. In practice, setting conditions to simplify the function $g(\theta)$ is much easier than setting conditions to simplify $\mathrm{Var}\{g(\theta)\}$. Besides, the second approach also provides the limiting distribution of the estimator, and this is very useful in practice. We will use this variance of linear approximation approach throughout this thesis.

1.7 Outline of thesis

The structure of this thesis is given as follows. Chapter 2 addresses the asymptotic properties of ML and REML estimators in the nested error regression model with balanced data ($m_i = m$) but without assuming normality, when both the number of clusters and cluster size go to infinity. Jiang (1996) addressed that asymptotic difference occurs when rank X increases with the sample size. However, he did not elaborate relationship between the number of cluster and clusters. We first distinguish a between cluster covariate that is constant within clusters $x_{ij} = x_i$ from a within cluster covariate that varies within clusters. This distinction is often ignored by using the form $\mu(x_{ij}) = \beta_0 + \beta_1 x_{ij}$ for both cases. Indeed, an important finding in this thesis is that this form does not work asymptotically when the m's grow for a within cluster covariate. The reason is that the information for within cluster parameters $\boldsymbol{\omega}^{(w)} = [\beta_2, \sigma_e^2]^T$ grows with $n = \sum_{i=1}^{g} m_i$ and the information for between cluster parameters $\boldsymbol{\omega}^{(b)} = [\beta_0, \beta_1, \sigma_\alpha^2]^T$ grows with g. We establish a uniform asymptotic linear approximation to the ML estimating equations that hold in an appropriate shrinking neighborhood of the true parameters, and then obtain consistency and asymptotic normality for the ML estimators from the uniform asymptotic linear approximation. We use the nice structure of balanced data (same cluster size) and use a similar approach to obtain the asymptotic properties of REML estimators when both of the number of clusters g and cluster sizes m go to infinity.

Chapter 3 extends the results of Chapter 2 to unbalanced data. This extension is by no means trivial because, in summing the squares of the deviations of the cluster means from the grand mean when constructing the ML and REML estimators, each cluster is given a different weight m_i equal to the number of the cluster size i.e., m_i.Nevertheless, we are able to establish a uniform asymptotic linear approximation to the ML estimating equations that hold in an appropriate shrinking neighborhood of the true parameters, and then obtain consistency and asymptotic normality for the ML estimators from the uniform asymptotic linear approximation. To establish the asymptotic properties of the REML estimator, we use a profile function argument instead of working directly with the structure of the estimating equation as we did in the balanced case. We firstly study an adjusted maximum likelihood (AML) estimating equation with the additional term $\log|\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}|$, and show that it has same asymptotic properties obtained from the log-likelihood with ML estimators. Then we use a profile function approach to show that the REML estimating equation is the profile version of the adjusted maximum likelihood estimating function. Finally, we obtain the asymptotic properties of REML estimators in general (non-normal) nested regression model with unbalanced data, when both the number of clusters and cluster size go to infinity.

Chapter 4 studies the asymptotic distribution of the EBLUP without assuming normality in linear mixed models. We use the asymptotic properties of ML and REML estimates given mild and reasonable conditions to directly prove that the asymptotic distribution of the EBLUP converges to the true distribution of the corresponding random effects. When the number of independent clusters grows, the ML and REML estimators of the between cluster parameters are consistent. When both the number of clusters and the cluster sizes tend to infinity, the ML and REML estimators of the within cluster parameters are also consistent. The consistency of ML and REML estimators is a vital step to study the asymptotic distribution of the EBLUP. We start with the balanced one-way model in approaching these developments, as it facilitates easier understanding and insight, before considering the nested error regression model. Under different rates of growth on the number of clusters and the cluster sizes, we obtain three different convolution distribution. Moreover, we derive a closed form approximation of the mean squared error of the EBLUP and compare it with other approximations from Kackar and Harville (1984) and Prasad and Rao (1990). Finally, we perform a Monte Carlo simulation with different settings. The simulation results indicate that as both the number of clusters and the cluster sizes grow, the average relative bias of our mean squared error is much more accurate under a normal distribution and mixture distribution.

Chapter 2 Balanced Data: Asymptotic Analysis of ML and REML Estimates for Hierarchical Linear Mixed Models with Large Cluster Sizes

2.1 Introduction

Regression models with nested errors are widely used in applied statistics to model relationships in clustered data, such as in small area estimation, for example, Battese et al. (1988), Rao (2015) and etc. The maximum likelihood (ML) or restricted maximum likelihood (REML) estimators usually used to fit these models are nonlinear, so asymptotic results provide an essential way to understand their properties and subsequently for constructing a basis for inference about the unknown parameters, make predictions, etc. The general asymptotic framework increases the number of clusters while keeping the size of each cluster bounded. However, in some applications, when the size of the clusters is also large, it is interesting to see what happens asymptotically when both the number of clusters and their sizes are allowed to grow. In particular, as we will show later on this thesis, this framework with increasing cluster size is needed to ensure consistent prediction.

Consider balanced data that (y_{ij}, x_{ij}) comprising of g independent clusters each with cluster size m, where y_{ij} is a scalar response variable, x_{ij} is a scalar explanatory variable or covariate. The nested error regression model specifies that

$$y_{ij} = \mu(x_{ij}) + \alpha_i + e_{ij}, \qquad j = 1, \dots, m, i = 1, \dots, g, \tag{2.1.1}$$

where $\mu(x_{ij})$ is the conditional mean of the response given the covariate x_{ij}, α_i is a random effect representing a random intercept or cluster effect, and e_{ij} is an error term. Note that this model is (1.5.1) again, but now with balanced data. We assume that the $\{\alpha_i\}$ and $\{e_{ij}\}$ are all mutually independent with mean zero and variances (also referred to in this thesis as the variance components) σ_α^2 and σ_e^2, respectively. We can estimate both variance components consistently when $g \to \infty$ with m bounded, but we need to let $m \to \infty$ to predict the random effects $\{\alpha_i\}$ consistently. If

$m \to \infty$ but g is held fixed, we can estimate the within cluster variance σ_e^2 consistently but not the between cluster variance σ_α^2. This motivates allowing both $g \to \infty$ and $m \to \infty$. This framework has surprising consequences for the conditional mean, which we parameterize as follows,

$$\mu(x_{ij}) = \begin{cases} \beta_0 + \beta_1 x_i & \text{if } x_{ij} = x_i \\ \beta_0 + \beta_1 \bar{x}_{i.} + \beta_2(x_{ij} - \bar{x}_{i.}) & \text{otherwise} \end{cases}, \tag{2.1.2}$$

where the βs are unknown coefficients, called regression parameters or fixed effects, and $\bar{x}_{i.} = m^{-1}\sum_{j=1}^m x_{ij}$ is the within cluster mean of the explanatory variable. In (2.1.2) we distinguish a between cluster covariate that is constant within clusters ($x_{ij} = x_i$) from a within cluster covariate that varies within clusters. This distinction is often ignored by using $\mu(x_{ij}) = \beta_0 + \beta_1 x_{ij}$ for both cases. An important finding in this thesis is that this form does not work asymptotically (when $m \to \infty$) for a within cluster covariate. The reason is that the information for within cluster parameters $\boldsymbol{\omega}^{(w)} = [\beta_2, \sigma_e^2]^T$ grows with $n = gm$, while the information for between cluster parameters $\boldsymbol{\omega}^{(b)} = [\beta_0, \beta_1, \sigma_\alpha^2]^T$ grows with g. The information for the coefficient corresponding to a within cluster covariate whose mean is non-zero grows with both n and g, making it difficult to normalize and control both sources of information at the same time. In contrast, if we partition the within cluster covariate as in (2.1.2) into a between cluster component and a pure within cluster component (which has all within cluster means zero), we obtain separate coefficients for which the information grows with n and g separately, and straightforward normalization is possible.

The model defined by (2.1.1) - (2.1.2) with two variance components and a single covariate is a straightforward example of a linear mixed model. Its simplicity allows us to make explicit calculations (involving only relatively low dimensional vectors and matrices) to see exactly what is going on and obtain precise results for asymptotic variances which we can interpret. Our calculations and results also show how to proceed to more complicated cases, with both multiple covariates and variance components. Nonetheless, from a practical point of view, our simple model is widely used in fields such as small area estimation to model and make predictions from clustered data, thus making our results immediately useful. Examples of small areas include a geographical region (e.g., a state, country, municipality, etc), a demographic group (e.g., a specific age $\times$ sex $\times$ race group), mean acreage under a crop for counties (small areas) in Iowa by Battese et al. (1988). Large clusters often occur in surveys of household poverty in administrative areas (Haslett, 2016).

In this chapter, the model (2.1.1) with balanced data (same cluster size) allows us to use Kronecker products, which simplifies derivations in many formulas and equations. Thus, we rewrite (2.1.1) in vector from as

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{1}_m\alpha_i + \mathbf{e}_i, \tag{2.1.3}$$

where $\mathbf{y}_i = (y_{i1}, \ldots, y_{im})^T$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^T$, $\mathbf{X}_i = [\mathbf{1}_m, \mathbf{1}_m\bar{x}_{i.}, \mathbf{x}_i - \mathbf{1}_m\bar{x}_{i.}]$ with $\mathbf{1}_m$ denoting an m-vector of ones and $\mathbf{x}_i = (x_{i1}, \ldots, x_{im})^T$, and $\mathbf{e}_i = (e_{i1}, \ldots, e_{im})^T$. Let $\mathbf{y}^T = (\mathbf{y}_1^T, \ldots, \mathbf{y}_g^T)$ and $\mathbf{x}^T = (\mathbf{x}_1^T, \ldots, \mathbf{x}_g^T)$. Under the model, we have

$$\mathrm{E}\,\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} = \mathbf{1}_m\beta_0 + \mathbf{1}_m\bar{x}_{i.}\beta_1 + (\mathbf{x}_i - \mathbf{1}_m\bar{x}_{i.})\beta_2, \qquad \mathrm{Var}(\mathbf{y}_i) = \sigma_\alpha^2\mathbf{J}_m + \sigma_e^2\mathbf{I}_m$$
$$\text{and} \qquad \mathrm{Var}(\mathbf{y}) = \mathbf{I}_g \otimes (\sigma_\alpha^2\mathbf{J}_m + \sigma_e^2\mathbf{I}_m).$$

where $\mathbf{J}_m = \mathbf{1}_m\mathbf{1}_m^T$. Let $\boldsymbol{\omega} = \mathrm{vec}\{\boldsymbol{\omega}^{(b)}, \boldsymbol{\omega}^{(w)}\} = (\beta_0, \beta_1, \sigma_\alpha^2, \beta_2, \sigma_e^2)^T$ and let $\boldsymbol{\theta} = (\sigma_\alpha^2, \sigma_e^2)^T$. Then we can write

$$\mathbf{V}_i^{-1}(\boldsymbol{\theta}) = \frac{1}{\sigma_e^2}\mathbf{I}_m - \frac{\sigma_\alpha^2}{\sigma_e^2(\sigma_e^2 + m\sigma_\alpha^2)}\mathbf{J}_m \qquad \text{and} \qquad \|\mathbf{V}(\boldsymbol{\theta})\| = (\sigma_e^2 + m\sigma_\alpha^2)^g\sigma_e^{2g(m-1)}.$$

To ease notation, we further denote

$$\bar{y}_{i.} = \frac{1}{m}\sum_{j=1}^{m} y_{ij}, \qquad \bar{y}_{..} = \frac{1}{gm}\sum_{i=1}^{g}\sum_{j=1}^{m} y_{ij}, \qquad \bar{x}_{..} = \frac{1}{gm}\sum_{i=1}^{g}\sum_{j=1}^{m} x_{ij},$$
$$S_{ijw}^{xy} = (x_{ij} - \bar{x}_{i.})(y_{ij} - \bar{y}_{i.}), \qquad S_{ib}^{xy} = (\bar{x}_{i.} - \bar{x}_{..})(\bar{y}_{i.} - \bar{y}_{..}),$$
$$S_w^{xy} = \sum_{i=1}^{g}\sum_{j=1}^{m} S_{ijw}^{xy} \qquad \text{and} \qquad S_b^{xy} = m\sum_{i=1}^{g} S_{ib}^{xy}.$$

To simplify notation, let $S_w^{xx} = S_w^x, S_w^{yy} = S_w^y, S_b^{xx} = S_b^x$ and $S_b^{yy} = S_b^y$. When the random effects α_i and e_{ij} in (2.1.1) are all normally distributed, it follows that the observations $\mathbf{y}_i$ are also normally distributed. We write the density of $\mathbf{y}$ given $\mathbf{x}$ as

$$L(\mathbf{y}; \boldsymbol{\omega}) = (2\pi)^{-\frac{n}{2}}|\mathbf{V}(\boldsymbol{\theta})|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{g}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^T\mathbf{V}_i^{-1}(\boldsymbol{\theta})(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})\right\}.$$

For the quadratic form in the log-likelihood, we have the following

$$
\begin{aligned}
&\sum_{i=1}^{g}(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})^T\mathbf{V}_i(\boldsymbol{\theta})(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})\\
&=\sum_{i=1}^{g}\{\mathbf{y}_i-\mathbf{1}_m\beta_0-\mathbf{1}_m\bar{x}_{i.}\beta_1-(\mathbf{x}_i-\mathbf{1}_m\bar{x}_{i.})\beta_2\}^T\mathbf{V}_i^{-1}(\boldsymbol{\theta})\{\mathbf{y}_i-\mathbf{1}_m\beta_0-\mathbf{1}_m\bar{x}_{i.}\beta_1-(\mathbf{x}_i-\mathbf{1}_m\bar{x}_{i.})\beta_2\}\\
&=\sum_{i=1}^{g}\{\mathbf{y}_i-\mathbf{1}_m\beta_0-\mathbf{1}_m\bar{x}_{i.}\beta_1-(\mathbf{x}_i-\mathbf{1}_m\bar{x}_{i.})\beta_2\}^T\left\{\frac{1}{\sigma_e^2}\mathbf{I}_m-\frac{\sigma_\alpha^2}{\sigma_e^2(\sigma_e^2+m\sigma_\alpha^2)}\mathbf{J}_m\right\}\{\mathbf{y}_i-\mathbf{1}_m\beta_0\\
&\qquad-\mathbf{1}_m\bar{x}_{i.}\beta_1-(\mathbf{x}_i-\mathbf{1}_m\bar{x}_{i.})\beta_2\}\\
&=\frac{1}{\sigma_e^2}\sum_{i=1}^{g}\{\mathbf{y}_i-\mathbf{1}_m\beta_0-\mathbf{1}_m\bar{x}_{i.}\beta_1-(\mathbf{x}_i-\mathbf{1}_m\bar{x}_{i.})\beta_2\}^T\{\mathbf{y}_i-\mathbf{1}_m\beta_0-\mathbf{1}_m\bar{x}_{i.}\beta_1-(\mathbf{x}_i-\mathbf{1}_m\bar{x}_{i.})\beta_2\}\\
&\qquad-\sum_{i=1}^{g}\frac{\sigma_\alpha^2}{\sigma_e^2(\sigma_e^2+m\sigma_\alpha^2)}\left[\sum_{j=1}^{m}\{y_{ij}-\beta_0-\bar{x}_{i.}\beta_1-(x_{ij}-\bar{x}_{i.})\beta_2\}\right]^2\\
&=\frac{1}{\sigma_e^2}\sum_{i=1}^{g}\sum_{j=1}^{m}(y_{ij}-\beta_0-\bar{x}_{i.}\beta_1-x_{ij}\beta_2+\bar{x}_{i.}\beta_2)^2-\sum_{i=1}^{g}\frac{\sigma_\alpha^2}{\sigma_e^2(\sigma_e^2+m\sigma_\alpha^2)}\left\{\sum_{j=1}^{m}(y_{ij}-\beta_0-\bar{x}_{i.}\beta_1-x_{ij}\beta_2+\bar{x}_{i.}\beta_2)\right\}^2\\
&=\frac{1}{\sigma_e^2}\sum_{i=1}^{g}\sum_{j=1}^{m}(y_{ij}-\beta_0-\bar{x}_{i.}\beta_1-x_{ij}\beta_2+\bar{x}_{i.}\beta_2)^2-\sum_{i=1}^{g}\frac{m^2\sigma_\alpha^2}{\sigma_e^2(\sigma_e^2+m\sigma_\alpha^2)}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1-\bar{x}_{i.}\beta_2+\bar{x}_{i.}\beta_2)^2\\
&=\frac{1}{\sigma_e^2}\sum_{i=1}^{g}\sum_{j=1}^{m}(y_{ij}-\beta_0-\bar{x}_{i.}\beta_1-x_{ij}\beta_2+\bar{x}_{i.}\beta_2)^2-\sum_{i=1}^{g}\frac{m}{\sigma_e^2}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1)^2+\sum_{i=1}^{g}\frac{m}{\sigma_e^2}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1)^2\\
&\qquad-\sum_{i=1}^{g}\frac{m^2\sigma_\alpha^2}{\sigma_e^2(\sigma_e^2+m\sigma_\alpha^2)}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1)^2\\
&=\frac{1}{\sigma_e^2}\sum_{i=1}^{g}\{(\mathbf{y}_i-\mathbf{1}_m\bar{y}_{i.})-(\mathbf{x}_i-\mathbf{1}_m\bar{x}_{i.})\beta_2\}^T\{(\mathbf{y}_i-\mathbf{1}_m\bar{y}_{i.})-(\mathbf{x}_i-\mathbf{1}_m\bar{x}_{i.})\beta_2\}+\sum_{i=1}^{g}\frac{m}{\sigma_e^2}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1)^2\\
&\qquad-\sum_{i=1}^{g}\frac{m^2\sigma_\alpha^2}{\sigma_e^2(\sigma_e^2+m\sigma_\alpha^2)}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1)^2\\
&=\frac{1}{\sigma_e^2}\sum_{i=1}^{g}\{(\mathbf{y}_i-\mathbf{1}_m\bar{y}_{i.})-(\mathbf{x}_i-\mathbf{1}_m\bar{x}_{i.})\beta_2\}^T\{(\mathbf{y}_i-\mathbf{1}_m\bar{y}_{i.})-(\mathbf{x}_i-\mathbf{1}_m\bar{x}_{i.})\beta_2\}+\sum_{i=1}^{g}\frac{m}{(\sigma_e^2+m\sigma_\alpha^2)}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1)^2\\
&=\frac{1}{\sigma_e^2}\sum_{i=1}^{g}\{(\mathbf{y}_i-\mathbf{1}_m\bar{y}_{i.})^T(\mathbf{y}_i-\mathbf{1}_m\bar{y}_{i.})-2(\mathbf{y}_i-\mathbf{1}_m\bar{y}_{i.})^T(\mathbf{x}_i-\mathbf{1}_m\bar{x}_{i.})\beta_2+(\mathbf{x}_i-\mathbf{1}_m\bar{x}_{i.})^T(\mathbf{x}_i-\mathbf{1}_m\bar{x}_{i.})\beta_2^2\}\\
&\qquad+\sum_{i=1}^{g}\frac{m}{(\sigma_e^2+m\sigma_\alpha^2)}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1)^2\\
&=\frac{1}{\sigma_e^2}(S_w^y-2S_w^{xy}\beta_2+S_w^x\beta_2^2)+\sum_{i=1}^{g}\frac{m}{(\sigma_e^2+m\sigma_\alpha^2)}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1)^2.
\end{aligned}
$$

The above results allow us to rewrite the density of function of $\mathbf{y}$ as

$$
L(\mathbf{y};\boldsymbol{\omega})=\frac{\exp\left[-\frac{1}{2}\left\{\frac{1}{\sigma_e^2}(S_w^y-2S_w^{xy}\beta_2+S_w^x\beta_2^2)+\sum_{i=1}^{g}\frac{m}{(\sigma_e^2+m\sigma_\alpha^2)}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1)^2\right\}\right]}{(2\pi)^{\frac{n}{2}}|\mathbf{V}(\boldsymbol{\theta})|^{\frac{1}{2}}}.
$$

Furthermore, the log-likelihood can be written as

$$
\begin{aligned}
l(\mathbf{y};\boldsymbol{\omega}) = &-\frac{n}{2}\log 2\pi - \frac{g}{2}\log(\sigma_e^2 + m\sigma_\alpha^2) - \frac{g(m-1)}{2}\log\sigma_e^2 - \frac{1}{2\sigma_e^2}(S_w^y - 2S_w^{xy}\beta_2 + S_w^x\beta_2^2) \\
&- \sum_{i=1}^{g}\frac{m}{2(\sigma_e^2 + m\sigma_\alpha^2)}(\bar{y}_{i.} - \beta_0 - \bar{x}_{i.}\beta_1)^2.
\end{aligned}
\tag{2.1.4}
$$

To maximize $l(\mathbf{y};\boldsymbol{\omega})$, we differentiate (2.1.4) with respect to $\beta_0, \beta_1, \beta_2, \sigma_\alpha^2$ and σ_e^2, respectively. We denote $l_{\beta_0}(\boldsymbol{\omega}) = \frac{\partial l(\mathbf{y};\boldsymbol{\omega})}{\partial\beta_0}$, $l_{\beta_1}(\boldsymbol{\omega}) = \frac{\partial l(\mathbf{y};\boldsymbol{\omega})}{\partial\beta_1}$, $l_{\sigma_\alpha^2}(\boldsymbol{\omega}) = \frac{\partial l(\mathbf{y};\boldsymbol{\omega})}{\partial\sigma_\alpha^2}$, $l_{\beta_2}(\boldsymbol{\omega}) = \frac{\partial l(\mathbf{y};\boldsymbol{\omega})}{\partial\beta_2}$ and $l_{\sigma_e^2}(\boldsymbol{\omega}) = \frac{\partial l(\mathbf{y};\boldsymbol{\omega})}{\partial\sigma_e^2}$. Let $\psi(\boldsymbol{\omega})^T = [\psi^{(b)}(\boldsymbol{\omega})^T, \psi^{(w)}(\boldsymbol{\omega})^T]$, where $\psi^{(b)}(\boldsymbol{\omega})^T = [l_{\beta_0}(\boldsymbol{\omega}), l_{\beta_1}(\boldsymbol{\omega}), l_{\sigma_\alpha^2}(\boldsymbol{\omega})]$ and $\psi^{(w)}(\boldsymbol{\omega})^T = [l_{\beta_2}(\boldsymbol{\omega}), l_{\sigma_e^2}(\boldsymbol{\omega})]$. We obtain the components of $\psi(\boldsymbol{\omega})$ as

$$
\begin{aligned}
l_{\beta_0}(\boldsymbol{\omega}) &= \sum_{i=1}^{g}\frac{m}{(\sigma_e^2 + m\sigma_\alpha^2)}(\bar{y}_{i.} - \beta_0 - \bar{x}_{i.}\beta_1) \\
l_{\beta_1}(\boldsymbol{\omega}) &= \sum_{i=1}^{g}\frac{m\bar{x}_{i.}}{(\sigma_e^2 + m\sigma_\alpha^2)}(\bar{y}_{i.} - \beta_0 - \bar{x}_{i.}\beta_1) \\
l_{\sigma_\alpha^2}(\boldsymbol{\omega}) &= -\frac{1}{2}\frac{gm}{(\sigma_e^2 + m\sigma_\alpha^2)} + \sum_{i=1}^{g}\frac{m^2}{2(\sigma_e^2 + m\sigma_\alpha^2)^2}(\bar{y}_{i.} - \beta_0 - \bar{x}_{i.}\beta_1)^2 \\
l_{\beta_2}(\boldsymbol{\omega}) &= \frac{1}{\sigma_e^2}S_w^{xy} - \frac{1}{\sigma_e^2}S_w^x\beta_2 \\
l_{\sigma_e^2}(\boldsymbol{\omega}) &= -\frac{g}{2(\sigma_e^2 + m\sigma_\alpha^2)} - \frac{g(m-1)}{2\sigma_e^2} + \frac{1}{2\sigma_e^4}(S_w^y - 2S_w^{xy}\beta_2 + S_w^x\beta_2^2) \\
&\quad + \sum_{i=1}^{g}\frac{m}{2(\sigma_e^2 + m\sigma_\alpha^2)^2}(\bar{y}_{i.} - \beta_0 - \bar{x}_{i.}\beta_1)^2.
\end{aligned}
\tag{2.1.5}
$$

To obtain estimating equation then solve the estimating equation $\psi(\boldsymbol{\omega}) = 0$. We consider the estimators obtained from the normal likelihood even when normality does not necessarily hold. Hartley and Rao (1967) proposed a computational algorithm for the solution of the likelihood equations and proved that under certain restrictions, the estimates were consistent and asymptotically normal as the size of the experimental design increased. Anderson (1969) proved that the estimates were consistent and asymptotically normal as the entire design was repeated. Miller (1977) studied asymptotic properties of the maximum likelihood estimates for a large class of design sequences whose size increases to infinity under normality. However, there are few asymptotic results of ML estimation without normality. On another hand, Thompson (1962) proposed that the restricted or residual maximum likelihood (REML) method is a way of estimating dispersion parameters associated with a linear model. Jiang (2017) introduced earlier studies on the asymptotic behavior of REML. Das (1979) first published research on asymptotic properties of REML estimators, but his previous work on REML asymptotic remained large unknow for fifteen years. Richardson and Welsh (1994) considered asymptotic properties of the REML estimator in hierarchical LMMs without normality by setting some conditions, including letting the number of clusters go to infinity, but all cluster sizes and the dimension of fixed effect were bounded. The results of Richardson and Welsh (1994) show that the REML estimator in the nested structure is consistent and asymptotically normal. Jiang (1996) studied the asymptotic behavior of the REML estimator without any assumption on the structure of the model, boundedness of the number of fixed effects and normality. Furthermore, he was the first person to also allow the number of fixed effects to tend to infinity. The results show that the REML estimator applied to data with a general structure (i.e., not only nested structure) is consistent and asymptotically normal as the number of fixed effects increase with the sample size. He considered dimensions of each random effect that can be viewed as a function of sample size n. When it applies in a nested structure, it is not a clear relationship between the number of clusters and cluster size. Richardson and Welsh (1994) figured out the increasing sample size without the boundedness of the number of clusters but fixed cluster sizes. Based on the above literature review, we notice that there remain some unanswered questions. For example, when sample size depends only on cluster size rather than the number of clusters, will the REML estimator still be consistent and asymptotically normal? When sample size depends on both the number of clusters and cluster size, how to ensure the REML estimators keep consistent? Answering such a question is not only important, but also forms the basis for exploring other as elements of inference. For example, Jiang (1998) proved that the empirical distribution of EBLUPs converges to the true distribution of the corresponding random effects. However, the results require further conditions beyond those needed to establish the consistency of the REML given in Jiang (1996). Additionally, for nested designs a further restriction on the cluster size increasing is necessary to ensure that the empirical distribution of EBLUPs converges to the true distribution of the corresponding random effects. Xie and Yang (2003) studied asymptotic properties for generalized estimating equations with large cluster size, but his results for M-estimators are not applicable to linear mixed models.

The difficulty with trying to apply general results to particular models like (2.1.1), is that it can be difficult to understand both the conditions (e.g. whether there is any restriction on the relationship between the cluster size and the number of clusters) and the main result. Main result itself is the case with the elegant general results of Jiang (1996). Although some nested structures satisfy his so-called invariant class AI^4 condition, this condition is complicated. Also, in furthermore work, increasing cluster size is a consequence of other complicated assumptions, while further conditions on the way the cluster size increases make it relatively difficult to understand whether cluster size relates to the number of clusters and leaves open questions of whether these conditions are minimal or not. To overcome these challenges, we approach this problem from the opposite direction by considering a particular, simple model (without assuming normality) and use elementary arguments to obtain simple, interpretable results as $g \to \infty$ and $m_L \to \infty$ for both maximum likelihood and REML estimators that give insight and build intuition into what we can expect in more general cases. Although simple, our two-component model (2.1.1) is already useful in practice, such as an example from Arora and Lahiri (1997), and we obtain interesting and unanticipated results including the importance of treating within cluster covariates carefully and the asymptotic orthogonality of between and within cluster parameters, even when the responses are not normally distributed. The arguments used to obtain these results are based on establishing a uniform asymptotic linear approximation to the estimating equations that holds in an appropriate shrinking neighborhood of the true parameters, and then obtaining consistency and asymptotic normality for the estimators from the uniform asymptotic linear approximation. The approach to establishing the uniform asymptotic linear approximation is inspired by the work of Bickel (1975) on one-step regression estimators; it has been applied to ML and REML estimators in linear mixed models by Richardson and Welsh (1994), Field et al. (2008), and Yoon and Welsh (2020). The arguments can be further extended to accommodate vector covariates and multiple error structures.

2.2 Maximum Likelihood Estimating Equation

Throughout this thesis, we assume the true model which describes the actual data generating mechanism is given by (2.1.1) and (2.1.2) with a true parameter $\boldsymbol{\omega}_0 = [\dot{\beta}_0, \dot{\beta}_1, \dot{\sigma_\alpha}^2, \dot{\beta}_2, \dot{\sigma_e}^2]^T$, and take all expectations under the true model. We use the subscript zero with a vector parameter and a dot above a scalar parameter to denote the true parameter value. With respect to the ordering of the parameters in $\boldsymbol{\omega}$ and $\boldsymbol{\omega}_0$, the between parameters and the within parameters are grouped is important. Let $\boldsymbol{\beta}_0 = [\dot{\beta}_0, \dot{\beta}_1, \dot{\beta}_2]^T$ and $\boldsymbol{\theta}_0 = [\dot{\sigma_\alpha}^2, \dot{\sigma_e}^2]^T$ denote the true regression parameters and variance components, respectively. We work mainly with the "impure" within cluster covariate case, i.e. x_{ij} varies within clusters and the within cluster means $\bar{x}_{i.}$ are not all equal, such that there three unknown regression parameters in (2.1.2) to estimate. If we have a pure within cluster covariate, then the constant cluster means are absorbed into the intercept and hence there is not β_1 to estimate, while if we have a between cluster covariate, there is no β_2 to estimate. The results for these cases can be obtained as special cases of the general results by deleting the components of vectors and the rows and columns of matrices corresponding to the excluded parameter.

To control the estimating function and derive the asymptotic properties of $\hat{\boldsymbol{\omega}}$ from the estimating equation, we impose the following conditions.

Condition A

1. The model (2.1.1) holds with true parameters $\boldsymbol{\omega}_0$ inside the parameter space Ω.

2. The number of clusters $g \to \infty$ and the cluster size $m \to \infty$.

3. The random variables $\{\alpha_i\}$ and $\{e_{ij}\}$ are independent and identically distributed, and there is a $\delta > 0$ such that $\mathrm{E}\,|\alpha_i|^{4+\delta} < \infty$ and $\mathrm{E}\,|e_{ij}|^{4+\delta} < \infty$ for all $i = 1, \dots, g$ and $j = 1, \dots, m$.

4. For the sequence of within cluster means $\{\bar{x}_{i.}\}$, there exist finite constants c_1 and c_2 such that $\lim_{g\to\infty} g^{-1}\sum_{i=1}^{g} \bar{x}_{i.}^{k} = c_k$, $k = 1, 2$ and a $\delta > 0$ such that $\lim_{g\to\infty} g^{-1}\sum_{i=1}^{g} |\bar{x}_{i.}|^{2+\delta} < \infty$. For the sequence $\{S_w^x\}$, there exists a constant c_3 such that $\lim_{g\to\infty}\lim_{m\to\infty} n^{-1}S_w^x = c_3 > 0$ and a $\delta > 0$ such that $\lim_{g\to\infty}\lim_{m\to\infty} n^{-1}\sum_{i=1}^{g}\sum_{j=1}^{m} |x_{ij} - \bar{x}_{i.}|^{2+\delta} < \infty$.

These are mild conditions which are often satisfied in practice. Conditions A3 and A4 ensure that limits needed to ensure the existence and finiteness of the asymptotic variance of the estimating function exist, and that we can establish a Lyapounov version of the central limit theorem for the estimating function. Condition A4 holds if the x_{ij} are realizations of clustered random variables that have finite $(4+\lambda)$ moments. In the balanced case, if the x_{ij} also have a similar cluster structure to the y_{ij} so $\mathrm{E}(x_{ij}) = \mu_x$, $\mathrm{Var}(x_{ij}) = \tau_x^2 + \sigma_x^2$, and the x_{ij}s in different clusters are independent while the x_{ij}s in the same cluster have the same covariance τ_x^2, then we have $c_1 = \mu_x$,$c_2 = \tau_x^2 + \mu_x^2$ and $c_3 = \sigma_x^2$.

For a pure within cluster covariate, we replace Condition A4 by

4'. For the sequence $\{x_{ij}\}$, there exists a finite constant c_3 such that $\lim_{g\to\infty}\lim_{m\to\infty} n^{-1}\sum_{i=1}^{g}\sum_{j=1}^{m_i} x_{ij}^2 = c_3$ and a $\delta > 0$ such that $\lim_{g\to\infty}\lim_{m\to\infty} n^{-1}\sum_{i=1}^{g}\sum_{j=1}^{m_i} |x_{ij}|^{2+\delta} < \infty$.

For a between cluster covariate $\{x_i\}$, we replace Condition A4 by

4". For the sequence $\{x_i\}$, there exist finite constants c_1 and c_2 such that $\lim_{g\to\infty} g^{-1}\sum_{i=1}^{g} x_i^k = c_k$, $k = 1, 2$ and a $\delta > 0$ such that $\lim_{g\to\infty} g^{-1}\sum_{i=1}^{g} |x_i|^{2+\delta} < \infty$.

For a sequence $\{U_n\}$ of random elements and a non-stochastic sequence $\{v_n\}$, we write $U_n = O_p(v_n)$ to mean U_n/v_n is bounded in probability and $U_n = o_p(v_n)$ to mean U_n/v_n converges to zero in probability.

2.2.1 Approximating the ML estimating equations evaluated at $\boldsymbol{\omega}_0$

The final central limit theorem is strongly driven by the behavior of the estimating equation evaluated at $\boldsymbol{\omega}_0$. We therefore start by developing a simple approximation to the estimating equation evaluated at $\boldsymbol{\omega}_0$. First, we write $\bar{y}_{i.} = \dot{\beta}_0 + \dot{\beta}_1\bar{x}_{i.} + \alpha_i + \bar{e}_{i.}$, so that

$$y_{ij} - \bar{y}_{i.} = (x_{ij} - \bar{x}_{i.})\dot{\beta}_2 + (e_{ij} - \bar{e}_{i.}) \qquad \text{and} \qquad \bar{y}_{i.} - \dot{\beta}_0 - \dot{\beta}_1\bar{x}_{i.} = \alpha_i + \bar{e}_{i.}$$

Then $\mathrm{Var}(\alpha_i) = \dot{\sigma_\alpha}^2$, $\mathrm{Var}(\bar{e}_{i.}) = \frac{\dot{\sigma_e}^2}{m}$, so

$$\bar{y}_{i.} - \dot{\beta}_0 - \dot{\beta}_1\bar{x}_{i.} = \alpha_i + o_p(1), \tag{2.2.1}$$

and

$$(\bar{y}_{i.} - \dot{\beta}_0 - \dot{\beta}_1\bar{x}_{i.})^2 = \alpha_i^2 + o_p(1). \tag{2.2.2}$$

Next, we obtain approximations for S_w^{xy} and S_w^y, respectively. For S_w^{xy}, we have

$$\begin{aligned} S_w^{xy} &= \sum_{i=1}^{g}\sum_{j=1}^{m}(x_{ij} - \bar{x}_{i.})(y_{ij} - \bar{y}_{i.}) \\ &= \sum_{i=1}^{g}\sum_{j=1}^{m}\Big\{(x_{ij} - \bar{x}_{i.})^2\dot{\beta}_2 + (x_{ij} - \bar{x}_{i.})(e_{ij} - \bar{e}_{i.})\Big\} \\ &= \sum_{i=1}^{g}\sum_{j=1}^{m}\Big\{(x_{ij} - \bar{x}_{i.})^2\dot{\beta}_2 - (x_{ij} - \bar{x}_{i.})e_{ij}\Big\} \\ &= S_w^x\dot{\beta}_2 + \sum_{i=1}^{g}\sum_{j=1}^{m}(x_{ij} - \bar{x}_{i.})e_{ij} \end{aligned} \tag{2.2.3}$$

and thus

$$\mathrm{E}(S_w^{xy}) = S_w^x\dot{\beta}_2. \tag{2.2.4}$$

Also,

$$\begin{aligned} \mathrm{Var}(S_w^{xy}) &= \mathrm{Var}\left[\sum_{i=1}^{g}\sum_{j=1}^{m}\Big\{(x_{ij} - \bar{x}_{i.})^2\dot{\beta}_2 - (x_{ij} - \bar{x}_{i.})e_{ij}\Big\}\right] \\ &= \sum_{i=1}^{g}\sum_{j=1}^{m}(x_{ij} - \bar{x}_{i.})^2\,\mathrm{Var}(e_{ij}) \\ &= \dot{\sigma_e}^2 S_w^x \\ &= O(gm). \end{aligned}$$

For S_w^y, we write

$$\begin{aligned} S_w^y &= \sum_{i=1}^{g}\sum_{j=1}^{m}(y_{ij}-\bar{y}_{i.})^2 \\ &= \sum_{i=1}^{g}\sum_{j=1}^{m}\left\{(x_{ij}-\bar{x}_{i.})\dot{\beta}_2+(e_{ij}-\bar{e}_{i.})\right\}^2 \\ &= \sum_{i=1}^{g}\sum_{j=1}^{m}\left\{(x_{ij}-\bar{x}_{i.})^2\dot{\beta}_2^2+2(x_{ij}-\bar{x}_{i.})\dot{\beta}_2(e_{ij}-\bar{e}_{i.})+(e_{ij}-\bar{e}_{i.})^2\right\} \\ &= S_w^x\dot{\beta}_2^2+2\dot{\beta}_2\sum_{i=1}^{g}\sum_{j=1}^{m}(x_{ij}-\bar{x}_{i.})e_{ij}+\sum_{i=1}^{g}\sum_{j=1}^{m}(e_{ij}^2-\dot{\sigma_e}^2)+mg\dot{\sigma_e}^2-m\sum_{i=1}^{g}\bar{e}_{i.}^2, \end{aligned} \tag{2.2.5}$$

so that

$$\mathrm{E}(S_w^y)=S_w^x\dot{\beta}_2^2+mg\dot{\sigma_e}^2-g\dot{\sigma_e}^2=S_w^x\dot{\beta}_2^2+g(m-1)\dot{\sigma_e}^2. \tag{2.2.6}$$

Furthermore, we calculate

$$\mathrm{Var}\left\{\dot{\beta}_2\sum_{i=1}^{g}\sum_{j=1}^{m}(x_{ij}-\bar{x}_{i.})e_{ij}\right\}=\dot{\beta}_2^2\sum_{i=1}^{g}\sum_{j=1}^{m}(x_{ij}-\bar{x}_{i.})^2\,\mathrm{Var}(e_{ij})=\dot{\beta}_2^2S_w^x\dot{\sigma_e}^2=O(gm),$$

$$\mathrm{Var}(\sum_{i=1}^{g}\sum_{j=1}^{m}e_{ij}^2)=\sum_{i=1}^{g}\sum_{j=1}^{m}\mathrm{Var}(e_{ij}^2)=\sum_{i=1}^{g}\sum_{j=1}^{m}\left\{\mathrm{E}(e_{ij}^4)-\dot{\sigma_e}^4\right\}=gm\left\{\mathrm{E}(e_{ij}^4)-\dot{\sigma_e}^4\right\}=O(gm)$$

and

$$\begin{aligned} \mathrm{Var}(m\bar{e}_{i.}^2) &= \mathrm{Var}\left\{m^{-1}(m\bar{e}_{i.})^2\right\} \\ &= m^{-2}\,\mathrm{Var}\left\{(\sum_{j=1}^{m}e_{ij})^2\right\} \\ &= m^{-2}\,\mathrm{Var}(\sum_{j=1}^{m}\sum_{k=1}^{m}e_{ij}e_{ik}) \\ &= m^{-2}\left\{\mathrm{E}(\sum_{j=1}^{m}\sum_{k=1}^{m}e_{ij}e_{ik})^2-\mathrm{E}^2(\sum_{j=1}^{m}\sum_{k=1}^{m}e_{ij}e_{ik})\right\} \\ &= m^{-2}\sum_{j=1}^{m}\sum_{k=1}^{m}\sum_{r=1}^{m}\sum_{s=1}^{m}\mathrm{E}(e_{ij}e_{ik}e_{ir}e_{is})-m^{-2}\left\{\sum_{j=1}^{m}\sum_{k=1}^{m}\mathrm{E}(e_{ij}e_{ik})\right\}^2 \\ &= m^{-2}\sum_{j=1}^{m}\mathrm{E}(e_{ij}^4)+(3-3m^{-1})\dot{\sigma_e}^4-\dot{\sigma_e}^4 \\ &= m^{-1}\,\mathrm{E}(e_{ij}^4)+(2-3m^{-1})\dot{\sigma_e}^4 \\ &= O(1). \end{aligned}$$

The key step here was a result given in [Cramér (1999), p345]. We write this result as Lemma 1 later. Combining the above results along with $\mathrm{Var}(S_w^y) = O(gm)$, then we can write

$$S_w^y = mg\dot{\sigma_e}^2 + S_w^x \dot{\beta}_2^2 + 2\dot{\beta}_2 \sum_{i=1}^{g}\sum_{j=1}^{m}(x_{ij} - \bar{x}_{i.})e_{ij} + \sum_{i=1}^{g}\sum_{j=1}^{m}(e_{ij}^2 - \dot{\sigma_e}^2) + o_p(g^{\frac{1}{2}}m^{\frac{1}{2}}). \quad (2.2.7)$$

Also, combining (2.2.3) and (2.2.7), we obtain

$$S_w^y - 2S_w^{xy}\dot{\beta}_2 + S_w^x\dot{\beta}_2^2 = gm\dot{\sigma_e}^2 + \sum_{i=1}^{g}\sum_{j=1}^{m}(e_{ij}^2 - \dot{\sigma_e}^2) + o_p(g^{\frac{1}{2}}m^{\frac{1}{2}}). \quad (2.2.8)$$

Finally, let $\mathbf{K} = \mathrm{diag}[g^{-\frac{1}{2}}\mathbf{I}_{(b)}, g^{-\frac{1}{2}}m^{-\frac{1}{2}}\mathbf{I}_{(w)}]$, where diag is diagonal matrix. Note that (b) and (w) have not been introduced as representing numbers. Just use words and say the number of between cluster parameters equals 3 and the number of within parameters equals 2. On combining (2.2.1), (2.2.2) and(2.2.8), we obtain

$$\mathbf{K}^{-\frac{1}{2}}\psi(\boldsymbol{\omega}_0) = \mathbf{K}^{-\frac{1}{2}}\xi(\boldsymbol{\omega}_0) + o_p(1), \quad (2.2.9)$$

where $\mathbf{K} = \mathrm{diag}[g^{-\frac{1}{2}}\mathbf{I}_{(b)}, g^{-\frac{1}{2}}m^{-\frac{1}{2}}\mathbf{I}_{(w)}]$. where $\xi(\boldsymbol{\omega}_0)$ has component,

$$\begin{aligned}
\xi_{\beta_0}(\boldsymbol{\omega}_0) &= \sum_{i=1}^{g} \frac{m}{(\dot{\sigma_e}^2 + m\dot{\sigma_\alpha}^2)}\alpha_i, \\
\xi_{\beta_1}(\boldsymbol{\omega}_0) &= \sum_{i=1}^{g} \frac{m}{(\dot{\sigma_e}^2 + m\dot{\sigma_\alpha}^2)}\bar{x}_{i.}\alpha_i, \\
\xi_{\sigma_\alpha^2}(\boldsymbol{\omega}_0) &= \sum_{i=1}^{g} \frac{m^2}{2(\dot{\sigma_e}^2 + m\dot{\sigma_\alpha}^2)^2}(\alpha_i^2 - \dot{\sigma_\alpha}^2), \\
\xi_{\beta_2}(\boldsymbol{\omega}_0) &= \frac{1}{\dot{\sigma_e}^2}\sum_{i=1}^{g}\sum_{j=1}^{m}(x_{ij} - \bar{x}_{i.})e_{ij} \qquad \text{and} \\
\xi_{\sigma_e^2}(\boldsymbol{\omega}_0) &= \frac{1}{2\dot{\sigma_e}^4}\sum_{i=1}^{g}\sum_{j=1}^{m}(e_{ij}^2 - \dot{\sigma_e}^2).
\end{aligned} \quad (2.2.10)$$

Again, we define

$$\xi(\boldsymbol{\omega}_0) = \begin{bmatrix} \xi^{(b)}(\boldsymbol{\omega}_0) \\ \xi^{(w)}(\boldsymbol{\omega}_0) \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{g} \xi_i^{(b)}(\boldsymbol{\omega}_0) \\ \sum_{i=1}^{g}\sum_{j=1}^{m} \xi_{ij}^{(w)}(\boldsymbol{\omega}_0) \end{bmatrix}. \quad (2.2.11)$$

2.2.2 Computing G_n, the variance of the approximate ML estimating equation at $\boldsymbol{\omega}_0$

Using (2.2.10), we calculate the variance of $\mathbf{K}^{-\frac{1}{2}}\xi(\boldsymbol{\omega}_0)$ as

$$\mathrm{Var}\left\{g^{-\frac{1}{2}}\xi_{\beta_0}(\boldsymbol{\omega}_0)\right\} = \frac{m^2\dot{\sigma_\alpha}^2}{(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)^2}, \qquad \mathrm{Cov}\left\{g^{-\frac{1}{2}}\xi_{\beta_0}(\boldsymbol{\omega}_0), g^{-\frac{1}{2}}\xi_{\beta_1}(\boldsymbol{\omega}_0)\right\} = \frac{1}{g}\sum_{i=1}^{g}\frac{m^2\dot{\sigma_\alpha}^2}{(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)^2}\bar{x}_{i.},$$
$$\mathrm{Cov}\left\{g^{-\frac{1}{2}}\xi_{\beta_0}(\boldsymbol{\omega}_0), g^{-\frac{1}{2}}\xi_{\dot{\sigma_\alpha}^2}(\boldsymbol{\omega}_0)\right\} = \frac{m^3}{2(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)^3}\,\mathrm{E}(\alpha_i^3), \mathrm{Var}\left\{g^{-\frac{1}{2}}\xi_{\beta_1}(\boldsymbol{\omega}_0)\right\} = \frac{1}{g}\sum_{i=1}^{g}\frac{m^2\dot{\sigma_\alpha}^2}{(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)^2}\bar{x}_{i.}^2,$$
$$\mathrm{Cov}\left\{g^{-\frac{1}{2}}\xi_{\beta_1}(\boldsymbol{\omega}_0), g^{-\frac{1}{2}}\xi_{\dot{\sigma_\alpha}^2}(\boldsymbol{\omega}_0)\right\} = \frac{1}{g}\sum_{i=1}^{g}\frac{m^3}{2(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)^3}\bar{x}_{i.}\,\mathrm{E}(\alpha_i^3),$$
$$\mathrm{Var}\left\{g^{-\frac{1}{2}}\xi_{\sigma_\alpha^2}(\boldsymbol{\omega}_0)\right\} = \frac{m^4}{4(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)^4}\left\{\mathrm{E}(\alpha_i^4)-\dot{\sigma_\alpha}^4\right\},$$
$$\mathrm{Var}\left\{g^{-\frac{1}{2}}m^{-\frac{1}{2}}\xi_{\beta_2}(\boldsymbol{\omega}_0)\right\} = \frac{S_w^x}{gm\dot{\sigma_e}^2}, \qquad \mathrm{Cov}\left\{g^{-\frac{1}{2}}m^{-\frac{1}{2}}\xi_{\beta_2}(\boldsymbol{\omega}_0), g^{-\frac{1}{2}}m^{-\frac{1}{2}}\xi_{\sigma_e^2}(\boldsymbol{\omega}_0)\right\} = 0$$
$$\text{and} \qquad \mathrm{Var}\left\{g^{-\frac{1}{2}}m^{-\frac{1}{2}}\xi_{\sigma_e^2}(\boldsymbol{\omega}_0)\right\} = \frac{1}{4\dot{\sigma_e}^8}(\mathrm{E}\,e_{ij}^4-\sigma_e^4).$$

From (2.2.10), we see that $\xi^{(b)}(\boldsymbol{\omega}_0)$ only depends on α_i and $\xi^{(w)}(\boldsymbol{\omega}_0)$ only depends on e_{ij}.

Since α_i and e_{ij} are independent, then we have

$$\mathbf{G}_n = \mathrm{Var}\begin{Bmatrix} g^{-\frac{1}{2}}\xi^{(b)}(\boldsymbol{\omega}_0) \\ g^{-\frac{1}{2}}m^{-\frac{1}{2}}\xi^{(w)}(\boldsymbol{\omega}_0)\end{Bmatrix} = \begin{bmatrix} \mathrm{Var}\left\{g^{-\frac{1}{2}}\xi^{(b)}(\boldsymbol{\omega}_0)\right\} & \mathbf{0} \\ \mathbf{0} & \mathrm{Var}\left\{g^{-\frac{1}{2}}m^{-\frac{1}{2}}\xi^{(w)}(\boldsymbol{\omega}_0)\right\}\end{bmatrix}$$
$$= \begin{bmatrix} \frac{m^2\dot{\sigma_\alpha}^2}{(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)^2} & \frac{1}{g}\sum_{i=1}^{g}\frac{m^2\dot{\sigma_\alpha}^2}{(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)^2}\bar{x}_{i.} & \frac{m^3}{2(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)^3}\,\mathrm{E}(\alpha_i^3) & 0 & 0 \\ \frac{1}{g}\sum_{i=1}^{g}\frac{m^2\dot{\sigma_\alpha}^2}{(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)^2}\bar{x}_{i.} & \frac{1}{g}\sum_{i=1}^{g}\frac{m^2\dot{\sigma_\alpha}^2}{(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)^2}\bar{x}_{i.}^2 & \frac{1}{g}\sum_{i=1}^{g}\frac{m^3}{2(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)^3}\bar{x}_{i.}\,\mathrm{E}(\alpha_i^3) & 0 & 0 \\ \frac{m^3}{2(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)^3}\,\mathrm{E}(\alpha_i^3) & \frac{1}{g}\sum_{i=1}^{g}\frac{m^3}{2(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)^3}\bar{x}_{i.}\,\mathrm{E}(\alpha_i^3) & \frac{m^4}{4(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)^4}\left\{\mathrm{E}(\alpha_i^4)-\dot{\sigma_\alpha}^4\right\} & 0 & 0 \\ 0 & 0 & 0 & \frac{S_w^x}{gm\dot{\sigma_e}^2} & 0 \\ 0 & 0 & 0 & 0 & \frac{\mathrm{E}(e_{ij}^4)-\sigma_e^4}{4\dot{\sigma_e}^8} \end{bmatrix}.$$

Since $\lim_{m\to\infty}\frac{m\dot{\sigma_\alpha}^2}{(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)} = 1$ and by Condition A3, we have

$$\begin{aligned}
&\lim_{g\to\infty}\lim_{m\to\infty}\frac{1}{g}\sum_{i=1}^{g}\frac{m^2\dot{\sigma_\alpha}^2}{(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)^2}\bar{x}_{i.} = \lim_{g\to\infty}\frac{\bar{x}_{i.}}{\dot{\sigma_\alpha}^2} = \frac{\mu_x}{\dot{\sigma_\alpha}^2},\\
&\lim_{g\to\infty}\lim_{m\to\infty}\frac{1}{g}\sum_{i=1}^{g}\frac{m^2\dot{\sigma_\alpha}^2}{(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)^2}\bar{x}_{i.}^2 = \lim_{g\to\infty}\frac{\bar{x}_{i.}^2}{\dot{\sigma_\alpha}^2} = \frac{\mu_x^2+\tau_x^2}{(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)^2}, \quad \text{and}\\
&\lim_{g\to\infty}\lim_{m\to\infty}\frac{1}{g}\sum_{i=1}^{g}\frac{m^3}{(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)^3}\bar{x}_{i.} = \lim_{g\to\infty}\frac{\bar{x}_{i.}}{\dot{\sigma_\alpha}^6} = \frac{\mu_x}{\dot{\sigma_\alpha}^6}.
\end{aligned} \tag{2.2.12}$$

Thus we have

$$\mathbf{G} = \lim_{g,m\to\infty} \mathbf{G}_n = \begin{bmatrix} \frac{1}{\dot{\sigma_\alpha}^2} & \frac{\mu_x}{\dot{\sigma_\alpha}^2} & \frac{\mathrm{E}(\alpha_i^3)}{2\dot{\sigma_\alpha}^6} & 0 & 0 \\ \frac{\mu_x}{\dot{\sigma_\alpha}^2} & \frac{\mu_x^2+\tau_x^2}{\dot{\sigma_\alpha}^2} & \frac{\mu_x\,\mathrm{E}(\alpha_i^3)}{2\dot{\sigma_\alpha}^6} & 0 & 0 \\ \frac{\mathrm{E}(\alpha_i^3)}{2\dot{\sigma_\alpha}^6} & \frac{\mu_x\,\mathrm{E}(\alpha_i^3)}{2\dot{\sigma_\alpha}^6} & \frac{\mathrm{E}\,\alpha_i^4-\dot{\sigma_\alpha}^4}{4\dot{\sigma_\alpha}^8} & 0 & 0 \\ 0 & 0 & 0 & \frac{\sigma_x^2}{\dot{\sigma_e}^2} & 0 \\ 0 & 0 & 0 & 0 & \frac{\mathrm{E}\,e_{ij}^4-\dot{\sigma_e}^4}{4\dot{\sigma_e}^8} \end{bmatrix} = \begin{bmatrix} \mathbf{G}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_2 \end{bmatrix}. \quad (2.2.13)$$

Here $\mathbf{G}_1$ is a 3×3 matrix and $\mathbf{G}_2$ is a 2×2 matrix . The elegant block diagonal form of $\mathbf{G}_n$ and $\mathbf{G}$ is a result of grouping the parameters into between and within cluster parameters. The more traditional grouping onto fixed and random parameters leads to less structured matrices and consequently less insight as we shall see below.

2.2.3 Computing the derivative matrix from the ML estimating equations

We need to compute the derivative and the expected derivative of the maximum likelihood estimating function $\boldsymbol{\psi}(\boldsymbol{\omega})$. We will apply the mean value theorem to each (real) element of $\boldsymbol{\psi}(\boldsymbol{\omega})$, and so we need to allow different arguments in each row of the derivative matrix. From (2.1.5), we write $\psi^{(b)}(\boldsymbol{\omega}) = \sum_{i=1}^{g} \psi_i^{(b)}(\boldsymbol{\omega})$ and $\psi^{(w)}(\boldsymbol{\omega}) = \sum_{i=1}^{g}\sum_{j=1}^{m} \psi_{ij}^{(w)}(\boldsymbol{\omega})$, where the components of $\psi_i^{(b)}(\boldsymbol{\omega})$ and $\psi_{ij}^{(w)}(\boldsymbol{\omega})$ are denoted as

$$\begin{aligned}
l_{i\beta_0}(\boldsymbol{\omega}) &= \frac{m}{(\sigma_e^2+m\sigma_\alpha^2)}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1) \\
l_{i\beta_1}(\boldsymbol{\omega}) &= \frac{m\bar{x}_{i.}}{(\sigma_e^2+m\sigma_\alpha^2)}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1) \\
l_{i\sigma_\alpha^2}(\boldsymbol{\omega}) &= -\frac{1}{2}\frac{m}{(\sigma_e^2+m\sigma_\alpha^2)} + \frac{m^2}{2(\sigma_e^2+m\sigma_\alpha^2)^2}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1)^2 \\
l_{ij\beta_2}(\boldsymbol{\omega}) &= \frac{1}{\sigma_e^2}S_{ijw}^{xy} - \frac{1}{\sigma_e^2}S_{ijw}^{x}\beta_2 \\
l_{ij\sigma_e^2}(\boldsymbol{\omega}) &= -\frac{1}{2m(\sigma_e^2+m\sigma_\alpha^2)} - \frac{m-1}{2m\sigma_e^2} + \frac{1}{2\sigma_e^4}(S_{ijw}^{y} - 2S_{ijw}^{xy}\beta_2 + S_{ijw}^{x}\beta_2^2) \\
&\quad + \frac{1}{2(\sigma_e^2+m\sigma_\alpha^2)^2}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1)^2.
\end{aligned}$$

Thus, we write

$$\nabla\psi(\boldsymbol{\omega}) = \begin{bmatrix} \nabla\psi^{(b)}(\boldsymbol{\omega}) \\ \nabla\psi^{(w)}(\boldsymbol{\omega}) \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{g} \nabla\psi_i^{(b)}(\boldsymbol{\omega}) \\ \sum_{i=1}^{g}\sum_{j=1}^{m_i} \nabla\psi_{ij}^{(w)}(\boldsymbol{\omega}) \end{bmatrix},$$

where

$$\nabla\psi_i^{(b)}(\boldsymbol{\omega}) = \begin{bmatrix} l_{i\beta_0\beta_0}(\boldsymbol{\omega}) & l_{i\beta_0\beta_1}(\boldsymbol{\omega}) & l_{i\beta_0\sigma_\alpha^2}(\boldsymbol{\omega}) & l_{i\beta_0\beta_2}(\boldsymbol{\omega}) & l_{i\beta_0\sigma_e^2}(\boldsymbol{\omega}) \\ l_{i\beta_1\beta_0}(\boldsymbol{\omega}) & l_{i\beta_1\beta_1}(\boldsymbol{\omega}) & l_{i\beta_1\sigma_\alpha^2}(\boldsymbol{\omega}) & l_{i\beta_1\beta_2}(\boldsymbol{\omega}) & l_{i\beta_1\sigma_e^2}(\boldsymbol{\omega}) \\ l_{i\sigma_e^2\beta_0}(\boldsymbol{\omega}) & l_{i\sigma_e^2\beta_1}(\boldsymbol{\omega}) & l_{i\sigma_e^2\sigma_\alpha^2}(\boldsymbol{\omega}) & l_{i\sigma_e^2\beta_2}(\boldsymbol{\omega}) & l_{i\sigma_e^2\sigma_e^2}(\boldsymbol{\omega}) \end{bmatrix},$$

and

$$\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega}) = \begin{bmatrix} l_{i\beta_2\beta_0}(\boldsymbol{\omega}) & l_{i\beta_2\beta_1}(\boldsymbol{\omega}) & l_{i\beta_2\sigma_\alpha^2}(\boldsymbol{\omega}) & l_{i\beta_2\beta_2}(\boldsymbol{\omega}) & l_{i\beta_2\sigma_e^2}(\boldsymbol{\omega}) \\ l_{i\sigma_e^2\beta_0}(\boldsymbol{\omega}) & l_{i\sigma_e^2\beta_1}(\boldsymbol{\omega}) & l_{i\sigma_e^2\sigma_\alpha^2}(\boldsymbol{\omega}) & l_{i\sigma_e^2\beta_2}(\boldsymbol{\omega}) & l_{i\sigma_e^2\sigma_e^2}(\boldsymbol{\omega}) \end{bmatrix}.$$

We compute each element of $\nabla\psi(\boldsymbol{\omega})$. For $\frac{\partial l_{i\beta_0}(\boldsymbol{\omega})}{\partial\boldsymbol{\omega}}$, the first row in $\nabla\psi_i^{(b)}(\boldsymbol{\omega})$, we have

$$\begin{aligned} &l_{i\beta_0\beta_0}(\boldsymbol{\omega}) = \frac{-m}{(\sigma_e^2+m\sigma_\alpha^2)}, \quad l_{i\beta_0\beta_1}(\boldsymbol{\omega}) = \frac{-m\bar{x}_{i.}}{(\sigma_e^2+m\sigma_\alpha^2)}, \quad l_{i\beta_0\sigma_\alpha^2}(\boldsymbol{\omega}) = \frac{-m^2}{(\sigma_e^2+m\sigma_\alpha^2)^2}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1), \\ &l_{i\beta_0\beta_2}(\boldsymbol{\omega}) = 0, \qquad \text{and} \quad l_{i\beta_0\sigma_e^2}(\boldsymbol{\omega}) = \frac{-m}{(\sigma_e^2+m\sigma_\alpha^2)^2}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1). \end{aligned} \tag{2.2.14}$$

For $\frac{\partial l_{i\beta_1}(\boldsymbol{\omega})}{\partial\boldsymbol{\omega}}$, the second row in $\nabla\psi_i^{(b)}(\boldsymbol{\omega})$, we have

$$\begin{aligned} &l_{i\beta_1\beta_0}(\boldsymbol{\omega}) = \frac{-m\bar{x}_{i.}}{(\sigma_e^2+m\sigma_\alpha^2)}, \qquad l_{i\beta_1\beta_1}(\boldsymbol{\omega}) = \frac{-m\bar{x}_{i.}^2}{(\sigma_e^2+m\sigma_\alpha^2)}, \qquad l_{i\beta_1\sigma_\alpha^2}(\boldsymbol{\omega}) = \frac{-m^2\bar{x}_{i.}}{(\sigma_e^2+m\sigma_\alpha^2)^2}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1), \\ &l_{i\beta_1\beta_2}(\boldsymbol{\omega}) = 0, \qquad \text{and} \quad l_{i\beta_1\sigma_e^2}(\boldsymbol{\omega}) = \frac{-m\bar{x}_{i.}}{(\sigma_e^2+m\sigma_\alpha^2)^2}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1). \end{aligned} \tag{2.2.15}$$

For $\frac{\partial l_{i\sigma_\alpha^2}(\boldsymbol{\omega})}{\partial\boldsymbol{\omega}}$,the third row in $\nabla\psi_i^{(b)}(\boldsymbol{\omega})$,we have

$$\begin{aligned} &l_{i\sigma_\alpha^2\beta_0}(\boldsymbol{\omega}) = \frac{-m^2}{(\sigma_e^2+m\sigma_\alpha^2)^2}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1), \qquad l_{i\sigma_\alpha^2\beta_1}(\boldsymbol{\omega}) = \frac{-m^2\bar{x}_{i.}}{(\sigma_e^2+m\sigma_\alpha^2)^2}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1), \\ &l_{i\sigma_\alpha^2\sigma_\alpha^2}(\boldsymbol{\omega}) = \frac{m^2}{2(\sigma_e^2+m\sigma_\alpha^2)^2} - \frac{m^3}{(\sigma_e^2+m\sigma_\alpha^2)^3}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1)^2, \\ &l_{i\sigma_\alpha^2\beta_2}(\boldsymbol{\omega}) = 0, \qquad \text{and} \qquad l_{i\sigma_\alpha^2\sigma_e^2}(\boldsymbol{\omega}) = \frac{m}{2(\sigma_e^2+m\sigma_\alpha^2)^2} - \frac{m^2}{(\sigma_e^2+m\sigma_\alpha^2)^3}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1)^2. \end{aligned} \tag{2.2.16}$$

For $\frac{\partial l_{ij\beta_2}(\boldsymbol{\omega})}{\partial\boldsymbol{\omega}}$, the first row in $\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega})$, we have

$$l_{ij\beta_2\beta_0}(\boldsymbol{\omega}) = l_{ij\beta_2\beta_1}(\boldsymbol{\omega}) = l_{ij\beta_2\sigma_\alpha^2}(\boldsymbol{\omega}) = 0, \quad l_{ij\beta_2\beta_2}(\boldsymbol{\omega}) = \frac{-1}{\sigma_e^2}S_{ijw}^{x}, \quad \text{and} \quad l_{ij\beta_2\sigma_e^2}(\boldsymbol{\omega}) = \frac{-1}{\sigma_e^4}(S_{ijw}^{xy} - S_{ijw}^{x}\beta_2). \tag{2.2.17}$$

Finally, for $\frac{\partial l_{ij\sigma_e^2}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}}$, the second row in$\nabla \psi_{ij}^{(w)}(\boldsymbol{\omega})$, we have

$$
\begin{aligned}
l_{ij\sigma_e^2\beta_0}(\boldsymbol{\omega}) &= \frac{-1}{(\sigma_e^2 + m\sigma_\alpha^2)^2}(\bar{y}_{i.} - \beta_0 - \bar{x}_{i.}\beta_1), \quad l_{ij\sigma_e^2\beta_1}(\boldsymbol{\omega}) = \frac{-\bar{x}_{i.}}{(\sigma_e^2 + m\sigma_\alpha^2)^2}(\bar{y}_{i.} - \beta_0 - \bar{x}_{i.}\beta_1), \\
l_{ij\sigma_e^2\sigma_\alpha^2}(\boldsymbol{\omega}) &= \frac{1}{2(\sigma_e^2 + m\sigma_\alpha^2)^2} - \frac{m}{(\sigma_e^2 + m\sigma_\alpha^2)^3}(\bar{y}_{i.} - \beta_0 - \bar{x}_{i.}\beta_1)^2, \\
l_{ij\sigma_e^2\beta_2}(\boldsymbol{\omega}) &= \frac{-1}{\sigma_e^4}(S_{ijw}^{xy} + S_{ijw}^{x}\beta_2), \quad \text{and} \\
l_{ij\sigma_e^2\sigma_e^2}(\boldsymbol{\omega}) &= \frac{1}{2m(\sigma_e^2 + m\sigma_\alpha^2)^2} + \frac{m-1}{2m\sigma_e^4} - \frac{1}{\sigma_e^6}(S_{ijw}^{y} - 2S_{ijw}^{xy}\beta_2 + S_{ijw}^{x}\beta_2^2) \\
&\quad - \frac{1}{(\sigma_e^2 + m\sigma_\alpha^2)^3}(\bar{y}_{i.} - \beta_0 - \bar{x}_{i.}\beta_1)^2.
\end{aligned}
\tag{2.2.18}
$$

We use the above terms to calculate the expected derivative matrix

$$
\mathrm{E}\left\{\nabla \psi(\boldsymbol{\omega})\right\} = \begin{bmatrix} \sum_{i=1}^{g} \mathrm{E}\left\{\nabla \psi_i^{(b)}(\boldsymbol{\omega})\right\} \\ \sum_{i=1}^{g}\sum_{j=1}^{m} \mathrm{E}\left\{\psi_{ij}^{(w)}(\boldsymbol{\omega})\right\} \end{bmatrix}.
$$

For $\mathrm{E}\left\{\frac{\partial l_{i\beta_0}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}}\right\}$, the first row of $\mathrm{E}\left\{\nabla \psi_i^{(b)}(\boldsymbol{\omega})\right\}$, we have

$$
\begin{aligned}
&\mathrm{E}\left\{l_{i\beta_0\beta_0}(\boldsymbol{\omega})\right\} = \frac{-m}{(\sigma_e^2 + m\sigma_\alpha^2)}, \quad \mathrm{E}\left\{l_{i\beta_0\beta_1}(\boldsymbol{\omega})\right\} = \frac{-m\bar{x}_{i.}}{(\sigma_e^2 + m\sigma_\alpha^2)} \\
&\mathrm{E}\left\{l_{i\beta_0\sigma_\alpha^2}(\boldsymbol{\omega})\right\} = \frac{-m^2}{(\sigma_e^2 + m\sigma_\alpha^2)^2}\left\{(\dot{\beta}_0 - \beta_0) + \bar{x}_{i.}(\dot{\beta}_2 - \beta_2)\right\}, \\
&\mathrm{E}\left\{l_{i\beta_0\beta_2}(\boldsymbol{\omega})\right\} = 0, \quad \text{and} \quad \mathrm{E}\left\{l_{i\beta_0\sigma_e^2}(\boldsymbol{\omega})\right\} = \frac{-m}{(\sigma_e^2 + m\sigma_\alpha^2)^2}\left\{(\dot{\beta}_0 - \beta_0) + \bar{x}_{i.}(\dot{\beta}_2 - \beta_2)\right\}.
\end{aligned}
\tag{2.2.19}
$$

Furthermore, when $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, we have

$$
\begin{aligned}
&\mathrm{E}\left\{l_{i\beta_0\beta_0}(\boldsymbol{\omega}_0)\right\} = \frac{-m}{(\dot{\sigma_e}^2 + m\dot{\sigma_\alpha}^2)}, \quad \mathrm{E}\left\{l_{i\beta_0\beta_1}(\boldsymbol{\omega}_0)\right\} = \frac{-m\bar{x}_{i.}}{(\dot{\sigma_e}^2 + m\dot{\sigma_\alpha}^2)}, \quad \mathrm{E}\left\{l_{i\beta_0\sigma_\alpha^2}(\boldsymbol{\omega}_0)\right\} = 0, \\
&\mathrm{E}\left\{l_{i\beta_0\beta_2}(\boldsymbol{\omega}_0)\right\} = 0 \quad \text{and} \quad \mathrm{E}\left\{l_{i\beta_0\sigma_e^2}(\boldsymbol{\omega}_0)\right\} = 0.
\end{aligned}
\tag{2.2.20}
$$

For $\mathrm{E}\left\{\frac{\partial l_{i\beta_1}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}}\right\}$, the second row of $\mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega})\right\}$, we calculate

$$\begin{aligned}
&\mathrm{E}\{l_{i\beta_1\beta_0}(\boldsymbol{\omega})\} = \frac{-m\bar{x}_{i.}}{(\sigma_e^2+m\sigma_\alpha^2)}, \qquad \mathrm{E}\{l_{i\beta_1\beta_1}(\boldsymbol{\omega})\} = \frac{-m\bar{x}_{i.}^2}{(\sigma_e^2+m\sigma_\alpha^2)},\\
&\mathrm{E}\left\{l_{i\beta_1\sigma_\alpha^2}(\boldsymbol{\omega})\right\} = \frac{-m^2\bar{x}_{i.}}{(\sigma_e^2+m\sigma_\alpha^2)^2}\left\{(\dot{\beta}_0-\beta_0)+\bar{x}_{i.}(\dot{\beta}_2-\beta_2)\right\},\\
&\mathrm{E}\{l_{i\beta_1\beta_2}(\boldsymbol{\omega})\} = 0, \qquad \text{and} \qquad \mathrm{E}\left\{l_{i\beta_1\sigma_e^2}(\boldsymbol{\omega})\right\} = \frac{-m\bar{x}_{i.}}{(\sigma_e^2+m\sigma_\alpha^2)^2}\left\{(\dot{\beta}_0-\beta_0)+\bar{x}_{i.}(\dot{\beta}_2-\beta_2)\right\}.
\end{aligned} \tag{2.2.21}$$

Also, when $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, we obtain

$$\begin{aligned}
&\mathrm{E}\{l_{i\beta_1\beta_0}(\boldsymbol{\omega}_0)\} = \frac{-m\bar{x}_{i.}}{(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)}, \qquad \mathrm{E}\{l_{i\beta_1\beta_1}(\boldsymbol{\omega}_0)\} = \frac{-m\bar{x}_{i.}^2}{(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)}, \qquad \mathrm{E}\left\{l_{i\beta_1\sigma_\alpha^2}(\boldsymbol{\omega}_0)\right\} = 0,\\
&\mathrm{E}\{l_{i\beta_1\beta_2}(\boldsymbol{\omega}_0)\} = 0, \qquad \text{and} \qquad \mathrm{E}\left\{l_{i\beta_1\sigma_e^2}(\boldsymbol{\omega}_0)\right\} = 0.
\end{aligned} \tag{2.2.22}$$

From (2.2.4) and (2.2.6), we have

$$\mathrm{E}\left\{(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1)^2\right\} = \left\{(\dot{\beta}_0-\beta_0)+\bar{x}_{i.}(\dot{\beta}_2-\beta_2)\right\}^2 + \dot{\sigma_\alpha}^2 + \frac{1}{m}\dot{\sigma_e}^2, \tag{2.2.23}$$

and

$$\mathrm{E}(S_{ijw}^x - 2S_{ijw}^{xy}\beta_2 + S_{ijw}^x\beta_2^2) = S_{ijw}^x(\dot{\beta}_1-\beta_1)^2 + \frac{m-1}{m}\dot{\sigma_e}^2. \tag{2.2.24}$$

For $\mathrm{E}\left\{\frac{\partial l_{i\sigma_\alpha^2}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}}\right\}$, the third row of $\mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega})\right\}$, using (2.2.23) and (2.2.24), we calculate

$$\begin{aligned}
&\mathrm{E}\left\{l_{i\sigma_\alpha^2\beta_0}(\boldsymbol{\omega})\right\} = \frac{-m^2}{(\sigma_e^2+m\sigma_\alpha^2)^2}\left\{(\dot{\beta}_0-\beta_0)+\bar{x}_{i.}(\dot{\beta}_2-\beta_2)\right\},\\
&\mathrm{E}\left\{l_{i\sigma_\alpha^2\beta_1}(\boldsymbol{\omega})\right\} = \frac{-m^2\bar{x}_{i.}}{(\sigma_e^2+m\sigma_\alpha^2)^2}\left\{(\dot{\beta}_0-\beta_0)+\bar{x}_{i.}(\dot{\beta}_2-\beta_2)\right\},\\
&\mathrm{E}\left\{l_{i\sigma_\alpha^2\sigma_\alpha^2}(\boldsymbol{\omega})\right\} = \frac{m^2}{2(\sigma_e^2+m\sigma_\alpha^2)^2} - \frac{m^3}{(\sigma_e^2+m\sigma_\alpha^2)^3}\left[\left\{(\dot{\beta}_0-\beta_0)+\bar{x}_{i.}(\dot{\beta}_2-\beta_2)\right\}^2 + \dot{\sigma_\alpha}^2 + m^{-1}\dot{\sigma_e}^2\right],\\
&\mathrm{E}\left\{l_{i\sigma_\alpha^2\beta_2}(\boldsymbol{\omega})\right\} = 0, \qquad \text{and}\\
&\mathrm{E}\left\{l_{i\sigma_\alpha^2\sigma_e^2}(\boldsymbol{\omega})\right\} = \frac{m}{2(\sigma_e^2+m\sigma_\alpha^2)^2} - \frac{m^2}{(\sigma_e^2+m\sigma_\alpha^2)^3}\left[\left\{(\dot{\beta}_0-\beta_0)+\bar{x}_{i.}(\dot{\beta}_2-\beta_2)\right\}^2 + \dot{\sigma_\alpha}^2 + m^{-1}\dot{\sigma_e}^2\right].
\end{aligned} \tag{2.2.25}$$

Furthermore, when $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, we have

$$\begin{aligned} &\mathrm{E}\left\{l_{i\sigma_\alpha^2\beta_0}(\boldsymbol{\omega}_0)\right\} = 0, \quad \mathrm{E}\left\{l_{i\sigma_\alpha^2\beta_1}(\boldsymbol{\omega}_0)\right\} = 0, \quad \mathrm{E}\left\{l_{i\sigma_\alpha^2\sigma_\alpha^2}(\boldsymbol{\omega}_0)\right\} = \frac{-m^2}{2(\dot{\sigma_e}^2 + m\dot{\sigma_\alpha}^2)^2}, \\ &\mathrm{E}\left\{l_{i\sigma_\alpha^2\beta_2}(\boldsymbol{\omega}_0)\right\} = 0, \quad \text{and} \quad \mathrm{E}\left\{l_{i\sigma_\alpha^2\sigma_e^2}(\boldsymbol{\omega}_0)\right\} = \frac{-m}{2(\dot{\sigma_e}^2 + m\dot{\sigma_\alpha}^2)^2}. \end{aligned} \tag{2.2.26}$$

Turning to $\mathrm{E}\left\{\frac{\partial l_{ij\beta_2}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}}\right\}$, the first row of $\mathrm{E}\left\{\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega})\right\}$, we calculate

$$\begin{aligned} &\mathrm{E}\left\{l_{ij\beta_2\beta_0}(\boldsymbol{\omega})\right\} = \mathrm{E}\left\{l_{ij\beta_2\beta_1}(\boldsymbol{\omega})\right\} = \mathrm{E}\left\{l_{ij\beta_2\sigma_\alpha^2}(\boldsymbol{\omega})\right\} = 0, \quad \mathrm{E}\left\{l_{ij\beta_2\beta_2}(\boldsymbol{\omega})\right\} = \frac{-1}{\sigma_e^2} S_{ijw}^x, \\ &\text{and} \quad \mathrm{E}\left\{l_{ij\beta_2\sigma_e^2}(\boldsymbol{\omega})\right\} = \frac{-1}{\sigma_e^4} S_{ijw}^x(\dot{\beta}_1 - \beta_1). \end{aligned} \tag{2.2.27}$$

Also, when $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, we obtain

$$\begin{aligned} &\mathrm{E}\left\{l_{ij\beta_2\beta_0}(\boldsymbol{\omega}_0)\right\} = \mathrm{E}\left\{l_{ij\beta_2\beta_1}(\boldsymbol{\omega}_0)\right\} = \mathrm{E}\left\{l_{ij\beta_2\sigma_\alpha^2}(\boldsymbol{\omega}_0)\right\} = 0, \quad \mathrm{E}\left\{l_{ij\beta_2\beta_2}(\boldsymbol{\omega}_0)\right\} = \frac{-1}{\dot{\sigma_e}^2} S_{ijw}^x, \\ &\text{and} \quad \mathrm{E}\left\{l_{ij\beta_2\sigma_e^2}(\boldsymbol{\omega}_0)\right\} = 0. \end{aligned} \tag{2.2.28}$$

For $\mathrm{E}\left\{\frac{\partial l_{ij\sigma_e^2}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}}\right\}$, the second row of $\mathrm{E}\left\{\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega})\right\}$, using (2.2.23) and (2.2.24), we have

$$\begin{aligned} \mathrm{E}\left\{l_{ij\sigma_e^2\beta_0}(\boldsymbol{\omega})\right\} &= \frac{-1}{(\sigma_e^2 + m\sigma_\alpha^2)^2}\left\{(\dot{\beta}_0 - \beta_0) + \bar{x}_{i.}(\dot{\beta}_2 - \beta_2)\right\}, \\ \mathrm{E}\left\{l_{ij\sigma_e^2\beta_1}(\boldsymbol{\omega})\right\} &= \frac{-\bar{x}_{i.}}{(\sigma_e^2 + m\sigma_\alpha^2)^2}\left\{(\dot{\beta}_0 - \beta_0) + \bar{x}_{i.}(\dot{\beta}_2 - \beta_2)\right\}, \\ \mathrm{E}\left\{l_{ij\sigma_e^2\sigma_\alpha^2}(\boldsymbol{\omega})\right\} &= \frac{1}{2(\sigma_e^2 + m\sigma_\alpha^2)^2} - \frac{m}{(\sigma_e^2 + m\sigma_\alpha^2)^3}\left[\left\{(\dot{\beta}_0 - \beta_0) + \bar{x}_{i.}(\dot{\beta}_2 - \beta_2)\right\}^2 + \dot{\sigma_\alpha}^2 + \frac{1}{m}\dot{\sigma_e}^2\right], \\ \mathrm{E}\left\{l_{ij\sigma_e^2\beta_2}(\boldsymbol{\omega})\right\} &= \frac{-1}{\sigma_e^4} S_{ijw}^x(\dot{\beta}_1 - \beta_1), \qquad \text{and} \\ \mathrm{E}\left\{l_{ij\sigma_e^2\sigma_e^2}(\boldsymbol{\omega})\right\} &= \frac{1}{2m(\sigma_e^2 + m\sigma_\alpha^2)^2} + \frac{m-1}{2m\sigma_e^4} - \frac{1}{\sigma_e^6}\left\{S_{ijw}^x(\dot{\beta}_1 - \beta_1)^2 + \frac{m-1}{m}\dot{\sigma_e}^2\right\} \\ &\quad - \frac{1}{(\sigma_e^2 + m\sigma_\alpha^2)^3}\left[\left\{(\dot{\beta}_0 - \beta_0) + \bar{x}_{i.}(\dot{\beta}_2 - \beta_2)\right\}^2 + \dot{\sigma_\alpha}^2 + \frac{1}{m}\dot{\sigma_e}^2\right]. \end{aligned} \tag{2.2.29}$$

When $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, we therefore obtain $\mathrm{E}\left\{\frac{\partial l_{ij\sigma_e^2}(\boldsymbol{\omega}_0)}{\partial \boldsymbol{\omega}}\right\}$

$$\mathrm{E}\left\{l_{ij\sigma_e^2\beta_0}(\boldsymbol{\omega}_0)\right\} = 0, \quad \mathrm{E}\left\{l_{ij\sigma_e^2\beta_1}(\boldsymbol{\omega}_0)\right\} = 0, \quad \mathrm{E}\left\{l_{ij\sigma_e^2\sigma_\alpha^2}(\boldsymbol{\omega}_0)\right\} = -\frac{1}{2(\dot{\sigma_e}^2 + m\dot{\sigma_\alpha}^2)^2},$$
$$\mathrm{E}\left\{l_{ij\sigma_e^2\beta_2}(\boldsymbol{\omega}_0)\right\} = 0, \quad \text{and} \quad \mathrm{E}\left\{l_{ij\sigma_e^2\sigma_e^2}(\boldsymbol{\omega}_0)\right\} = \frac{-1}{2m(\dot{\sigma_e}^2 + m\dot{\sigma_\alpha}^2)^2} - \frac{m-1}{2m\dot{\sigma_e}^4}. \tag{2.2.30}$$

Combining the results of (2.2.20), (2.2.22), (2.2.26), (2.2.28) and (2.2.30), we have

$$\mathbf{B}_n = -\mathbf{K}^{-\frac{1}{2}}\,\mathrm{E}\left\{\nabla\psi(\boldsymbol{\omega}_0)\right\}\mathbf{K}^{-\frac{1}{2}}$$
$$= \begin{bmatrix} \frac{m}{(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)} & \frac{1}{g}\sum_{i=1}^{g}\frac{m\bar{x}_{i.}}{(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)} & 0 & 0 & 0 \\ \frac{1}{g}\sum_{i=1}^{g}\frac{m\bar{x}_{i.}}{(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)} & \frac{1}{g}\sum_{i=1}^{g}\frac{m\bar{x}_{i.}^2}{(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)} & 0 & 0 & 0 \\ 0 & 0 & \frac{m^2}{2(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)^2} & 0 & \frac{m}{2(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)^2} \\ 0 & 0 & 0 & \frac{S_w^x}{gm\dot{\sigma_e}^2} & 0 \\ 0 & 0 & \frac{1}{2(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)^2} & 0 & \frac{1}{2(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)^2} + \frac{(m-1)}{2m\dot{\sigma_e}^4} \end{bmatrix}.$$

Furthermore, from applying the results of (2.2.12), we obtain

$$\mathbf{B} = \lim_{g,m\to\infty}\mathbf{B}_n = \begin{bmatrix} \frac{1}{\dot{\sigma_\alpha}^2} & \frac{\mu_x}{\dot{\sigma_\alpha}^2} & 0 & 0 & 0 \\ \frac{\mu_x}{\dot{\sigma_\alpha}^2} & \frac{\mu_x^2+\tau_x^2}{\dot{\sigma_\alpha}^2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2\dot{\sigma_\alpha}^4} & 0 & 0 \\ 0 & 0 & 0 & \frac{\sigma_x^2}{\dot{\sigma_e}^2} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2\dot{\sigma_e}^4} \end{bmatrix} = \begin{bmatrix} \mathbf{B}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_2 \end{bmatrix}. \tag{2.2.31}$$

Here $\mathbf{B}_1$ is a 3×3 matrix and $\mathbf{B}_2$ is a 2×2 matrix . From (2.2.13) and (2.2.31), we can see that under normality, it holds that $\mathbf{G} = \mathbf{B}$.

2.3 Main results on ML estimators

The first step in establishing a central limit theorem for the maximum likelihood estimator $\hat{\boldsymbol{\omega}}$ is to find an appropriate normalizing matrix $\mathbf{K}$ that ensures that $\mathbf{K}^{-\frac{1}{2}}\{\psi(\boldsymbol{\omega}_0) - \xi(\boldsymbol{\omega}_0)\} = o_p(1)$ and that $\mathbf{K}^{-\frac{1}{2}}\xi(\boldsymbol{\omega}_0) \xrightarrow{D} N(\mathbf{0}, \mathbf{G})$, where $\mathbf{G}$ is a positive definite covariance matrix. That is, $\mathbf{K}^{-\frac{1}{2}}\psi(\boldsymbol{\omega}_0)$ can be approximated by $\mathbf{K}^{-\frac{1}{2}}\xi(\boldsymbol{\omega}_0)$ which converges in distribution to a $N(\mathbf{0}, \mathbf{G})$ distribution. Under condition A, we show that the appropriate normalizing matrix is the diagonal matrix $\mathbf{K} = \mathrm{diag}[g, g, g, n, n]$. This justifies focusing our attention on the neighborhood of $\boldsymbol{\omega}_0$ defined by $\mathcal{B} = \{\boldsymbol{\omega} : |\mathbf{K}^{\frac{1}{2}}(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega}_0)| \le M\}$, where M is a finite constant. The second step is to show that, uniformly on $\mathcal{B}$, the derivative matrix $\nabla\psi(\boldsymbol{\omega})$ of $\psi(\boldsymbol{\omega})$, pre- and post-multiplied by $\mathbf{K}^{-\frac{1}{2}}$, converges in probability to a positive definite matrix $-\mathbf{B}$. We can then establish that, uniformly on $\mathcal{B}$,

$$\mathbf{K}^{-\frac{1}{2}}\psi(\boldsymbol{\omega}) = \mathbf{K}^{-\frac{1}{2}}\xi(\boldsymbol{\omega}) - \mathbf{B}\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0) + o_p(1).$$

This uniform asymptotic linearity result allows us to deduce the existence of a consistent root to the estimating equation that satisfies the central limit theorem with asymptotic variance.

Theorem 1. *Suppose Condition A holds. Then there exists a solution $\hat{\boldsymbol{\omega}}$ to the estimating equations $\psi(\boldsymbol{\omega}) = 0$, satisfying $|\mathbf{K}^{\frac{1}{2}}(\hat{\boldsymbol{\omega}}-\boldsymbol{\omega}_0)| = O_p(1)$ as $g \to \infty$ and $m \to \infty$, where $\mathbf{K} = \operatorname{diag}[g\mathbf{I}_{(b)}, mg\mathbf{I}_{(w)}]$. Moreover,*

$$\mathbf{K}^{\frac{1}{2}}(\hat{\boldsymbol{\omega}}-\boldsymbol{\omega}_0) = \mathbf{B}^{-1}\mathbf{K}^{-\frac{1}{2}}\xi(\boldsymbol{\omega}_0) + o_p(1),$$

whence

$$\mathbf{K}^{\frac{1}{2}}(\hat{\boldsymbol{\omega}}-\boldsymbol{\omega}_0) \xrightarrow{D}, N(\mathbf{0}, \mathbf{B}^{-1}\mathbf{G}\mathbf{B}^{-1}),$$

where $\mathbf{B} = \lim_{g,m\to\infty}\mathbf{B}_n = -\mathbf{K}^{-\frac{1}{2}}\operatorname{E}\{\nabla\psi(\boldsymbol{\omega}_0)\}\mathbf{K}^{-\frac{1}{2}}$ *and* $\mathbf{G} = \lim_{g,m\to\infty}\mathbf{G}_n = \operatorname{Var}\left\{\mathbf{K}^{-\frac{1}{2}}\xi(\boldsymbol{\omega}_0)\right\}$.

To simplify the proofs, we establish a preliminary lemma and then break the main proof into a sequence of lemmas.

Lemma 1 (Cramér (1999), p345)**.** *Suppose $e_{ij} \overset{i.i.d}{\sim} (0, \dot{\sigma_e}^2)$ with $\operatorname{E} e_{ij}^4 < \infty$. Then we have $\sum_{j,k,r,s=1}^{m}\operatorname{E}(e_{ij}e_{ik}e_{ir}e_{is}) = m\operatorname{E}(e_{ij}^4) + 3m(m-1)\dot{\sigma_e}^4$.*

Proof. See appendix. □

Lemma 2. *Suppose $\mathbf{u}$ is a stochastic $m \times 1$ vector. Then $\operatorname{E}|\mathbf{u}-\operatorname{E}\mathbf{u}|^2 \le \operatorname{E}|\mathbf{u}|^2 = \operatorname{tr}\left\{\operatorname{E}(\mathbf{u}^T\mathbf{u})\right\}$.*

Proof. We have

$$\operatorname{E}|\mathbf{u}-\operatorname{E}\mathbf{u}|^2 = \operatorname{E}\left\{\sum_{j=1}^{m}(u_j - \operatorname{E}u_i)^2\right\} = \sum_{j=1}^{m}\operatorname{Var}(u_j) \le \sum_{j=1}^{m}\operatorname{E}(u_j^2) = \operatorname{E}\left\{\sum_{j=1}^{m}(u_j)^2\right\} = \operatorname{E}|\mathbf{u}|^2.$$

□

We establish a central limit theorem for the maximum likelihood estimator $\hat{\boldsymbol{\omega}}$ is to find that $\mathbf{K}^{-\frac{1}{2}}\xi(\boldsymbol{\omega}_0) \xrightarrow{D} N(\mathbf{0}, \mathbf{G})$, where $\mathbf{G}$ is a positive definite covariance matrix.

Lemma 3. *Suppose Condition A holds. Then if $g \to \infty$ and $m \to \infty$, we have $\mathbf{G}_n^{-\frac{1}{2}}\mathbf{K}^{-\frac{1}{2}}\xi(\boldsymbol{\omega}_0) \xrightarrow{D} N(\mathbf{0}, \mathbf{I})$.*

Proof. The first step is to prove $g^{-\frac{1}{2}}\mathbf{G}_1^{-\frac{1}{2}}\xi^{(b)}(\boldsymbol{\omega}_0) \xrightarrow{D} N(\mathbf{0},\mathbf{I}_{(b)})$, where $\mathbf{I}_{(b)}$ is a 3×3 identity matrix. Suppose $\mathbf{a}$ is a fixed 3-vector satisfying $\mathbf{a}^T\mathbf{a}=1$. From the equation (2.2.11), we have $\mathbf{a}^T g^{-\frac{1}{2}}\mathbf{G}_1^{-\frac{1}{2}}\xi_1^{(b)}(\boldsymbol{\omega}_0),\ldots,\mathbf{a}^T g^{-\frac{1}{2}}\mathbf{G}_1^{-\frac{1}{2}}\xi_g^{(b)}(\boldsymbol{\omega}_0)$ are independent scalar random variables. As $g,m\to\infty$, then we can obtain $\mathbf{a}^T g^{-\frac{1}{2}}\mathbf{G}_1^{-\frac{1}{2}}\,\mathrm{E}[\xi_i^{(b)}(\boldsymbol{\omega}_0)]=0$ and $\mathrm{Var}[\mathbf{a}^T g^{-\frac{1}{2}}\mathbf{G}_1^{-\frac{1}{2}}\xi^{(b)}(\boldsymbol{\omega}_0)]$ $=\sum_{i=1}^{g}\mathrm{Var}[\mathbf{a}^T g^{-\frac{1}{2}}\mathbf{G}_1^{-\frac{1}{2}}\xi_i^{(b)}(\boldsymbol{\omega}_0)]=1$. There exists a $\delta>0$ such that

$$
\begin{aligned}
&\lim_{g\to\infty}\sum_{i=1}^{g}\mathrm{E}\,|\mathbf{a}^T g^{-\frac{1}{2}}\mathbf{G}_1^{-\frac{1}{2}}\xi_i^{(b)}(\boldsymbol{\omega}_0)|^{2+\delta}\\
&\le \lim_{g\to\infty} g^{-1-\frac{\delta}{2}}\|\mathbf{G}_1^{-\frac{1}{2}}\|^{2+\delta}\sum_{i=1}^{g}\mathrm{E}\,|\xi_i^{(b)}(\boldsymbol{\omega}_0)|^{2(1+\frac{\delta}{2})}\\
&= \lim_{g\to\infty} g^{-1-\frac{\delta}{2}}\|\mathbf{G}_1^{-\frac{1}{2}}\|^{2+\delta}\sum_{i=1}^{g}\mathrm{E}[\{\xi_{i\beta_0}^2(\boldsymbol{\omega}_0)+\xi_{i\beta_1}^2(\boldsymbol{\omega}_0)+\xi_{i\sigma_\alpha^2}^2(\boldsymbol{\omega}_0)\}^{1+\frac{\delta}{2}}]\\
&\le \lim_{g\to\infty} g^{-1-\frac{\delta}{2}}\|\mathbf{G}_1^{-\frac{1}{2}}\|^{2+\delta}\sum_{i=1}^{g}[\mathrm{E}\{\xi_{i\beta_0}^{2+\delta}(\boldsymbol{\omega}_0)\}+\mathrm{E}\{\xi_{i\beta_1}^{2+\delta}(\boldsymbol{\omega}_0)\}+\mathrm{E}\{\xi_{i\sigma_\alpha^2}^{2+\delta}(\boldsymbol{\omega}_0)\}]
\end{aligned}
$$

From (2.2.10), we have

$$
\begin{aligned}
\xi_{i\beta_0}^{2+\delta}(\boldsymbol{\omega}_0) &= \left\{\frac{m}{(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)}\right\}^{2+\delta}\alpha_i^{2+\delta},\\
\xi_{i\beta_1}^{2+\delta}(\boldsymbol{\omega}_0) &= \left\{\frac{m}{(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)}\bar{x}_{i.}\right\}^{2+\delta}\alpha_i^{2+\delta}\qquad\text{and}\\
\xi_{i\sigma_\alpha^2}^{2+\delta}(\boldsymbol{\omega}_0) &= \left\{\frac{m^2}{2(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)^2}\right\}^{2+\delta}(\alpha_i^2-\dot{\sigma_\alpha}^2)^{2+\delta}.
\end{aligned}
$$

Since $\mathrm{E}(\alpha_i)^{4+\lambda}<\infty$, and $\frac{m}{(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)}<\infty$, and also by Condition A3, we have $\mathrm{E}\left\{\xi_{i\beta_0}^{2+\delta}(\boldsymbol{\omega}_0)\right\}+\mathrm{E}\left\{\xi_{i\beta_1}^{2+\delta}(\boldsymbol{\omega}_0)\right\}+\mathrm{E}\left\{\xi_{i\sigma_\alpha^2}^{2+\delta}(\boldsymbol{\omega}_0)\right\}<\infty$. Thus, we obtain

$$
\lim_{g\to\infty} g^{-1-\frac{\delta}{2}}\|\mathbf{G}_1^{-\frac{1}{2}}\|^{2+\delta}\sum_{i=1}^{g}\left[\mathrm{E}\left\{\xi_{i\beta_0}^{2+\delta}(\boldsymbol{\omega}_0)\right\}+\mathrm{E}\left\{\xi_{i\beta_1}^{2+\delta}(\boldsymbol{\omega}_0)\right\}+\mathrm{E}\left\{\xi_{i\sigma_\alpha^2}^{2+\delta}(\boldsymbol{\omega}_0)\right\}\right]=O_p(g^{-\frac{\delta}{2}})=o_p(1).
$$

That is Lyapunov's condition [p.362, Billingsley (2008)] holds. Consequently $\mathbf{a}^T g^{-\frac{1}{2}}\mathbf{G}_1^{-\frac{1}{2}}\xi^{(b)}(\boldsymbol{\omega}_0)$ converges in distribution $N(0,1)$, as $g\to\infty$. The result follows from the Cramer-Wold device [Billingsley (1968) page 49].

The second step is to prove $g^{-\frac{1}{2}}m^{-\frac{1}{2}}\mathbf{G}_2^{-\frac{1}{2}}\xi^{(w)}(\boldsymbol{\omega}_0) \xrightarrow{D} N(\mathbf{0},\mathbf{I}_{(w)})$, where $\mathbf{I}_{(w)}$ is a 2×2 identity matrix. Suppose $\mathbf{b}$ is a fixed 2-vector with $\mathbf{b}^T\mathbf{b}=1$. From the equation (2.2.11), we have that $\mathbf{b}^T g^{-\frac{1}{2}}m^{-\frac{1}{2}}\mathbf{G}_2^{-\frac{1}{2}}\xi_{11}^{(w)}(\boldsymbol{\omega}_0),\ldots,\mathbf{b}^T g^{-\frac{1}{2}}m^{-\frac{1}{2}}\mathbf{G}_2^{-\frac{1}{2}}\xi_{gm}^{(w)}(\boldsymbol{\omega}_0)$ are independent scalar random variables. As $g,m\to\infty$, then $\mathbf{b}^T g^{-\frac{1}{2}}m^{-\frac{1}{2}}\mathbf{G}_2^{-\frac{1}{2}}\,\mathrm{E}[\xi_{ij}^{(w)}(\boldsymbol{\omega}_0)]=0$ and $\mathrm{Var}[\mathbf{b}^T g^{-\frac{1}{2}}m^{-\frac{1}{2}}\mathbf{G}_2^{-\frac{1}{2}}\xi^{(w)}(\boldsymbol{\omega}_0)]=$

$\sum_{i=1}^{g}\sum_{j=1}^{m}\mathrm{Var}[\mathbf{b}^T g^{-\frac{1}{2}} m^{-\frac{1}{2}} \mathbf{G}_2^{-\frac{1}{2}} \xi_{ij}^{(w)}(\boldsymbol{\omega}_0)] = 1$. Moreover, there exists a $\delta > 0$ such that

$$
\begin{aligned}
&\lim_{g\to\infty}\lim_{m\to\infty}\sum_{i=1}^{g}\sum_{j=1}^{m}\mathrm{E}\,|\mathbf{b}^T g^{-\frac{1}{2}} m^{-\frac{1}{2}} \mathbf{G}_2^{-\frac{1}{2}} \xi_{ij}^{(w)}(\boldsymbol{\omega}_0)|^{2+\delta} \\
&\le \lim_{g\to\infty}\lim_{m\to\infty} g^{-\frac{\delta}{2}} m^{-\frac{\delta}{2}} \|\mathbf{G}_2^{-\frac{1}{2}}\|^{2+\delta}\sum_{i=1}^{g}\sum_{j=1}^{m}\mathrm{E}\,|\xi_{ij}^{(w)}(\boldsymbol{\omega}_0)|^{2(1+\frac{\delta}{2})} \\
&\le \lim_{g\to\infty}\lim_{m\to\infty} g^{-\frac{\delta}{2}} m^{-\frac{\delta}{2}} \|\mathbf{G}_2^{-\frac{1}{2}}\|^{2+\delta}\sum_{i=1}^{g}\sum_{j=1}^{m}\mathrm{E}\left[\left\{\xi_{ij\beta_2}^{2}(\boldsymbol{\omega}_0) + \xi_{ij\sigma_e^2}^{2}(\boldsymbol{\omega}_0)\right\}^{1+\frac{\delta}{2}}\right] \\
&\le \lim_{g\to\infty}\lim_{m\to\infty} g^{-\frac{\delta}{2}} m^{-\frac{\delta}{2}} \|\mathbf{G}_2^{-\frac{1}{2}}\|^{2+\delta}\sum_{i=1}^{g}\sum_{j=1}^{m}\left[\mathrm{E}\left\{\xi_{ij\beta_2}^{2+\delta}(\boldsymbol{\omega}_0)\right\} + \mathrm{E}\left\{\xi_{ij\sigma_e^2}^{2+\delta}(\boldsymbol{\omega}_0)\right\}\right].
\end{aligned}
$$

Using (2.2.10) we calculate

$$
\xi_{ij\beta_2}^{2+\delta}(\boldsymbol{\omega}_0) = \left\{\frac{1}{\dot{\sigma_e}^2}(x_{ij} - \bar{x}_{i.})\right\}^{2+\delta} e_{ij}^{2+\delta} \qquad \text{and} \qquad \xi_{ij\sigma_e^2}^{2}(\boldsymbol{\omega}_0) = (\frac{1}{2\dot{\sigma_e}^4})^{2+\delta}(e_{ij}^2 - \dot{\sigma_e}^2)^{2+\delta}.
$$

Since $\mathrm{E}(e_{ij})^{4+\lambda} < \infty$ and under Condition A3, we obtain $\mathrm{E}\left\{\xi_{ij\beta_2}^{2+\delta}(\boldsymbol{\omega}_0)\right\} + \mathrm{E}\left\{\xi_{ij\sigma_e^2}^{2+\delta}(\boldsymbol{\omega}_0)\right\} < \infty$. Thus

$$
\lim_{g\to\infty}\lim_{m\to\infty} g^{-\frac{\delta}{2}} m^{-\frac{\delta}{2}} \|\mathbf{G}_2^{-\frac{1}{2}}\|^{2+\delta}\sum_{i=1}^{g}\sum_{j=1}^{m}\left[\mathrm{E}\left\{\xi_{ij\beta_2}^{2+\delta}(\boldsymbol{\omega}_0)\right\} + \mathrm{E}\left\{\xi_{ij\sigma_e^2}^{2+\delta}(\boldsymbol{\omega}_0)\right\}\right] = O_p(g^{-\frac{\delta}{2}} m^{-\frac{\delta}{2}}) = o_p(1).
$$

That is Lyapunov's condition holds. Consequently $\mathbf{b}^T g^{-\frac{1}{2}} m^{-\frac{1}{2}} \mathbf{G}_2^{-\frac{1}{2}} \xi^{(w)}$ converges in distribution $N(0,1)$, as $g \to \infty$ and $m \to \infty$. The result follows from the Cramer-Wold device. Since $\xi^{(b)}(\boldsymbol{\omega}_0)$ and $\xi^{(w)}(\boldsymbol{\omega}_0)$ are independent from (2.2.10), then we have $\mathbf{G}_n^{-\frac{1}{2}}\mathbf{K}^{-\frac{1}{2}}\xi(\boldsymbol{\omega}_0) \xrightarrow{D} N(\mathbf{0}, \mathbf{I})$. □

We prove preliminary lemmas and then the asymptotic linearity result.

Lemma 4. *Suppose Condition A holds, then, for $i = 1, \ldots, g$ and $j = 1, \ldots, m$,*

$$
\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} \mathrm{E}\,\|\nabla\psi_i^{(b)}(\boldsymbol{\omega})\|^2 < \infty \qquad \textit{and} \qquad \sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} \mathrm{E}\,\|\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega})\|^2 < \infty.
$$

Proof. From (2.2.14)-(2.2.18), and using Condition A3, we can find that

$$
\nabla\psi_i^{(b)}(\boldsymbol{\omega}) = \begin{bmatrix} O_p(1) & O_p(1) & O_p(1) & 0 & O_p(m^{-1}) \\ O_p(1) & O_p(1) & O_p(1) & 0 & O_p(m^{-1}) \\ O_p(1) & O_p(1) & O_p(1) & 0 & O_p(m^{-1}) \end{bmatrix}, \tag{2.3.1}
$$

and

$$\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega}) = \begin{bmatrix} 0 & 0 & 0 & O_p(1) & O_p(1) \\ O_p(m^{-2}) & O_p(m^{-2}) & O_p(m^{-2}) & O_p(1) & O_p(1) \end{bmatrix}.$$

Therefore we have $\|\nabla\psi_i^{(b)}(\boldsymbol{\omega})\|^2 = O_p(1)$ and $\|\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega})\|^2 = O_p(1)$. Hence

$$\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} \mathrm{E}\,\|\nabla\psi_i^{(b)}(\boldsymbol{\omega})\|^2 < \infty, \quad \text{and} \quad \sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} \mathrm{E}\,\|\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega})\|^2 < \infty.$$

$\square$

Lemma 5. *Suppose Condition A holds. Let $t_k = (t_{k1}, t_{k2}, t_{k3}, t_{k4}, t_{k5})^T$. Decompose the cube $\mathcal{B} = \{\boldsymbol{\omega} : |\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega} - \boldsymbol{\omega}_0)| \le M\}$ into smaller cubes $\mathcal{C} = \{\mathcal{C}(t_k)\}$, where M is a constant and $\boldsymbol{\omega} \in \mathcal{C}(t_k) = \{\boldsymbol{\omega} : |\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega} - \boldsymbol{\omega}_0)| \le Mg^{-\frac{1}{4}}\}$. As $g, m \to \infty$, then*

$$\max_{1\le k\le N}\sup_{\boldsymbol{\omega}\in\mathcal{C}(t_k)} g^{-\frac{1}{2}}\sum_{i=1}^{g}\left|\left[\nabla\psi_i^{(b)}(\tilde{\boldsymbol{\omega}}) - \mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega}^*)\right\}\right](\boldsymbol{\omega} - t_k)\right| = o_p(1),$$
$$\max_{1\le k\le N}\sup_{\boldsymbol{\omega}\in\mathcal{C}(t_k)} g^{-\frac{1}{2}}m^{-\frac{1}{2}}\sum_{i=1}^{g}\sum_{j=1}^{m}\left|\left[\nabla\psi_{ij}^{(w)}(\tilde{\boldsymbol{\omega}}) - \mathrm{E}\left\{\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega}^*)\right\}\right](\boldsymbol{\omega} - t_k)\right| = o_p(1),$$

where $N = g^{\frac{1}{4}}$ cubes and $\tilde{\boldsymbol{\omega}}, \boldsymbol{\omega}^* \in \mathcal{C}(t_k)$.

Proof. Let $\|\mathbf{K}^{-\frac{1}{2}}\| \le g^{-\frac{1}{2}}L$ for some constant $L > 0$. For the first equation in lemma 5, we have

$$\begin{aligned}
&\max_{1\le k\le N}\sup_{\boldsymbol{\omega}\in\mathcal{C}(t_k)} g^{-\frac{1}{2}}\sum_{i=1}^{g}\left|\left[\nabla\psi_i^{(b)}(\tilde{\boldsymbol{\omega}}) - \mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega}^*)\right\}\right](\boldsymbol{\omega} - t_k)\right| \\
&\le \max_{1\le k\le N}\sup_{\boldsymbol{\omega}\in\mathcal{C}(t_k)} g^{-\frac{1}{2}}\sum_{i=1}^{g}\|\nabla\psi_i^{(b)}(\tilde{\boldsymbol{\omega}}) - \mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega}^*)\right\}\|\|\mathbf{K}^{-\frac{1}{2}}\||\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega} - t_k)| \\
&\le \max_{1\le k\le N}\sup_{\boldsymbol{\omega}\in\mathcal{C}(t_k)} g^{-\frac{1}{4}}M\, g^{-\frac{1}{2}}\sum_{i=1}^{g}\|\nabla\psi_i^{(b)}(\tilde{\boldsymbol{\omega}}) - \mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega}^*)\right\}\|\|\mathbf{K}^{-\frac{1}{2}}\| \\
&\le \max_{1\le k\le N}\sup_{\boldsymbol{\omega}\in\mathcal{C}(t_k)} g^{-1-\frac{1}{4}}ML\sum_{i=1}^{g}\left\{\|\nabla\psi_i^{(b)}(\tilde{\boldsymbol{\omega}})\| + \mathrm{E}\,\|\nabla\psi_i^{(b)}(\boldsymbol{\omega}^*)\|\right\} \\
&= O_p(g^{-\frac{1}{4}}) = o_p(1),
\end{aligned}$$

where the last line follows from $\|\nabla\psi_i^{(b)}(\tilde{\boldsymbol{\omega}})\| = O_p(1)$ and $\mathrm{E}\,\|\nabla\psi_i^{(b)}(\boldsymbol{\omega}^*)\| = O_p(1)$ from (2.3.1).

For the second equation in lemma 5, we require a more delicate argument.

$$
\begin{aligned}
&\max_{1\le k\le N}\sup_{\boldsymbol{\omega}\in\mathcal{C}(t_k)} g^{-\frac{1}{2}}m^{-\frac{1}{2}}\sum_{i=1}^{g}\sum_{j=1}^{m}\left|\left[\nabla\psi_{ij}^{(w)}(\tilde{\boldsymbol{\omega}})-\mathrm{E}\left\{\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega}^*)\right\}\right](\boldsymbol{\omega}-t_k)\right| \\
&\le \max_{1\le k\le n}\sup_{\boldsymbol{\omega}\in\mathcal{C}(t_k)} g^{-\frac{1}{2}}m^{-\frac{1}{2}}\sum_{i=1}^{g}\sum_{j=1}^{m}\Bigg[\left|(\boldsymbol{\omega}-t_k)^T\left\{\frac{\partial l_{ij\beta_2}(\tilde{\boldsymbol{\omega}})}{\partial\boldsymbol{\omega}}-\mathrm{E}\,\frac{\partial l_{ij\beta_2}(\boldsymbol{\omega}^*)}{\partial\boldsymbol{\omega}}\right\}\right| \\
&\quad+\left|(\boldsymbol{\omega}-t_k)^T\left\{\frac{\partial l_{ij\sigma_e^2}(\tilde{\boldsymbol{\omega}})}{\partial\boldsymbol{\omega}}-\mathrm{E}\,\frac{\partial l_{ij\sigma_e^2}(\boldsymbol{\omega}^*)}{\partial\boldsymbol{\omega}}\right\}\right|\Bigg] \\
&\le \max_{1\le k\le n}\sup_{\boldsymbol{\omega}\in\mathcal{C}(t_k)} g^{-\frac{1}{2}}m^{-\frac{1}{2}}\sum_{i=1}^{g}\sum_{j=1}^{m}\big[\left|\left\{l_{ij\beta_2\beta_2}(\tilde{\boldsymbol{\omega}})-\mathrm{E}\,l_{ij\beta_2\beta_2}(\boldsymbol{\omega}^*)\right\}(\beta_2-t_{k4})\right| \\
&\quad+\left|\left\{l_{ij\beta_2\sigma_e^2}(\tilde{\boldsymbol{\omega}})-\mathrm{E}\,l_{ij\beta_2\sigma_e^2}(\boldsymbol{\omega}^*)\right\}(\sigma_e^2-t_{k5})\right| \\
&\quad+\left|\left\{l_{ij\sigma_e^2\beta_0}(\tilde{\boldsymbol{\omega}})-\mathrm{E}\,l_{ij\sigma_e^2\beta_0}(\boldsymbol{\omega}^*)\right\}(\beta_0-t_{k1})\right|+\left|\left\{l_{ij\sigma_e^2\beta_1}(\tilde{\boldsymbol{\omega}})-\mathrm{E}\,l_{ij\sigma_e^2\beta_1}(\boldsymbol{\omega}^*)\right\}(\beta_1-t_{k2})\right| \\
&\quad+\left|\left\{l_{ij\sigma_e^2\sigma_\alpha^2}(\tilde{\boldsymbol{\omega}})-\mathrm{E}\,l_{ij\sigma_e^2\sigma_\alpha^2}(\boldsymbol{\omega}^*)\right\}(\sigma_\alpha^2-t_{k3})\right|+\left|\left\{l_{ij\sigma_e^2\beta_2}(\tilde{\boldsymbol{\omega}})-\mathrm{E}\,l_{ij\sigma_e^2\beta_2}(\boldsymbol{\omega}^*)\right\}(\beta_2-t_{k4})\right| \\
&\quad+\left|\left\{l_{ij\sigma_e^2\sigma_e^2}(\tilde{\boldsymbol{\omega}})-\mathrm{E}\,l_{ij\sigma_e^2\sigma_e^2}(\boldsymbol{\omega}^*)\right\}(\sigma_e^2-t_{k5})\right|\big],
\end{aligned}
$$

where the last line follows from (2.2.17), (2.2.18), (2.2.27) and (2.2.29). We calculate

$$
|l_{ij\sigma_e^2\beta_0}(\tilde{\boldsymbol{\omega}})-\mathrm{E}\,l_{ij\sigma_e^2\beta_0}(\boldsymbol{\omega}^*)| = \left|\frac{-1}{(\tilde{\sigma}_e^2+m\tilde{\sigma}_\alpha^2)^2}(\bar{y}_{i.}-\tilde{\beta}_0-\bar{x}_{i.}\tilde{\beta}_1)-0\right| = \frac{1}{m^2}\left|\frac{\bar{y}_{i.}-\tilde{\beta}_0-\bar{x}_{i.}\tilde{\beta}_1}{(\tilde{\sigma}_\alpha^2+\frac{\tilde{\sigma}_e^2}{m})^2}\right|.
$$

From Condition A3, we know $\max_{1\le i\le g}|\bar{x}_{i.}|<\infty$, then we obtain $|l_{ij\sigma_e^2\beta_0}(\tilde{\boldsymbol{\omega}})-\mathrm{E}\,l_{ij\sigma_e^2\beta_0}(\boldsymbol{\omega}^*)| = O_p(m^{-2})$. Similarly, we have $|l_{ij\sigma_e^2\beta_1}(\tilde{\boldsymbol{\omega}})-\mathrm{E}\,l_{ij\sigma_e^2\beta_1}(\boldsymbol{\omega}^*)| = O_p(m^{-2})$ and $|l_{ij\sigma_e^2\sigma_\alpha^2}(\tilde{\boldsymbol{\omega}})-\mathrm{E}\,l_{ij\sigma_e^2\sigma_\alpha^2}(\boldsymbol{\omega}^*)| = O_p(m^{-2})$ as $m\to\infty$. Since $\boldsymbol{\omega}\in\mathcal{C}(t_k)$, then we have $|\beta_0-t_{k1}|\le g^{-\frac{3}{4}}M$, $|\beta_1-t_{k2}|\le g^{-\frac{3}{4}}M$, $|\sigma_\alpha^2-t_{k3}|\le g^{-\frac{3}{4}}M$,$|\beta_2-t_{k4}|\le g^{-\frac{3}{4}}m^{-\frac{1}{2}}M$ and $|\sigma_e^2-t_{k5}|\le g^{-\frac{3}{4}}m^{-\frac{1}{2}}M$. Thus, we obtain

$$
\max_{1\le k\le N}\sup_{\boldsymbol{\omega}\in\mathcal{C}(t_k)} g^{-\frac{1}{2}}m^{-\frac{1}{2}}\sum_{i=1}^{g}\sum_{j=1}^{m}\left|\left[\nabla\psi_{ij}^{(w)}(\tilde{\boldsymbol{\omega}})-\mathrm{E}\left\{\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega}^*)\right\}\right](\boldsymbol{\omega}-t_k)\right| = O(g^{-\frac{1}{4}}) = o_p(1).
$$

□

Lemma 6. *Suppose Condition A holds, then we have*

$$
\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} g^{-\frac{1}{2}}\sum_{i=1}^{g}\left|\left[\mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega})\right\}-\mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega}_0)\right\}\right](\boldsymbol{\omega}-\boldsymbol{\omega}_0)\right| = o(1),
$$

$$
\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} g^{-\frac{1}{2}}m^{-\frac{1}{2}}\sum_{i=1}^{g}\sum_{j=1}^{m}\left|\left[\mathrm{E}\left\{\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega})\right\}-\mathrm{E}\,\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega}_0)\right](\boldsymbol{\omega}-\boldsymbol{\omega}_0)\right| = o(1).
$$

Proof. For the first equation of lemma 6, we have

$$
\begin{aligned}
&\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} g^{-\frac{1}{2}}\sum_{i=1}^{g}|\left[\mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega})\right\}-\mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega}_0)\right\}\right](\boldsymbol{\omega}-\boldsymbol{\omega}_0)| \\
\le &\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} g^{-\frac{1}{2}}\sum_{i=1}^{g}\|\left[\mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega})\right\}-\mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega}_0)\right\}\right]\|\|\mathbf{K}^{-\frac{1}{2}}\|\|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)| \\
\le &\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} g^{-1}LM\sum_{i=1}^{g}\|\left[\mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega})\right\}-\mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega}_0)\right\}\right]\|,
\end{aligned}
$$

where $\|\mathbf{K}^{-\frac{1}{2}}\| \le g^{-\frac{1}{2}}L$ for some constant $L > 0$. For convenient calculation, we denote $\zeta = \frac{m}{(\sigma_e^2+m\sigma_\alpha^2)}$, $\dot\zeta = \frac{m}{(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)}$ and $h_i(\beta_0,\beta_1) = (\dot\beta_0-\beta_0)+\bar{x}_{i.}(\dot\beta_1-\beta_1)$. Since $|\mathbf{K}^{-\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)| \le M$ and Condition A3 implies that $\bar{x}_{i.}$ and $\bar{x}_{i.}^2$ are bounded as $g, m\to\infty$, then we have $|\beta_0-\dot\beta_0| \le g^{-\frac{1}{2}}M$, $|\beta_1-\dot\beta_1| \le g^{-\frac{1}{2}}M$, $|\sigma_\alpha^2-\dot{\sigma_\alpha}^2| \le g^{-\frac{1}{2}}M$, $|\beta_2-\dot\beta_2| \le g^{-\frac{1}{2}}m^{-\frac{1}{2}}M$ and $|\sigma_e^2-\dot{\sigma_e}^2| \le g^{-\frac{1}{2}}m^{-\frac{1}{2}}M$. Combining these results, we obtain

$$
\begin{aligned}
&|h_i(\beta_0,\beta_1)| = |(\dot\beta_0-\beta_0)+\bar{x}_{i.}(\dot\beta_1-\beta_1)| \le g^{-\frac{1}{2}}M, \\
&|\zeta-\dot\zeta| = |\frac{m}{(\sigma_e^2+m\sigma_\alpha^2)}-\frac{m}{(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)}| = |m\frac{(\dot{\sigma_e}^2-\sigma_e^2)+m(\dot{\sigma_\alpha}^2-\sigma_\alpha^2)}{(\sigma_e^2+m\sigma_\alpha^2)(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)}| \le g^{-\frac{1}{2}}M \\
&|\frac{\zeta}{\dot\zeta}-1| = |\frac{(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)}{(\sigma_e^2+m\sigma_\alpha^2)}-1| = |\frac{(\dot{\sigma_e}^2-\sigma_e^2)+m(\dot{\sigma_\alpha}^2-\sigma_\alpha^2)}{(\sigma_e^2+m\sigma_\alpha^2)}| \le g^{-\frac{1}{2}}M.
\end{aligned}
\tag{2.3.2}
$$

Based on (2.2.19), (2.2.20), (2.2.21), (2.2.22), (2.2.25) and (2.2.26), we can proceed to calculate $\|\mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega})-\nabla\psi_i^{(b)}(\boldsymbol{\omega}_0)\right\}\|$,

$$
|\mathrm{E}\{l_{i\beta_0\beta_2}(\boldsymbol{\omega})-l_{i\beta_0\beta_2}(\boldsymbol{\omega}_0)\}| = |\mathrm{E}\{l_{i\beta_1\beta_2}(\boldsymbol{\omega})-\mathrm{E}\,l_{i\beta_1\beta_2}(\boldsymbol{\omega}_0)\}| = |\mathrm{E}\left\{l_{i\sigma_\alpha^2\beta_2}(\boldsymbol{\omega})-l_{i\sigma_\alpha^2\beta_2}(\boldsymbol{\omega}_0)\right\}| = 0,
$$

$$
\begin{aligned}
&|\mathrm{E}\{l_{i\beta_0\beta_0}(\boldsymbol{\omega})-l_{i\beta_0\beta_0}(\boldsymbol{\omega}_0)\}| = |\zeta-\dot\zeta| \le g^{-\frac{1}{2}}M, \\
&|\mathrm{E}\{l_{i\beta_0\beta_1}(\boldsymbol{\omega})-l_{i\beta_0\beta_1}(\boldsymbol{\omega}_0)\}| = |(\zeta-\dot\zeta)\bar{x}_{i.}| \le g^{-\frac{1}{2}}M, \\
&|\mathrm{E}\left\{l_{i\beta_0\sigma_\alpha^2}(\boldsymbol{\omega})-l_{i\beta_0\sigma_\alpha^2}(\boldsymbol{\omega}_0)\right\}| = |\zeta h_i(\beta_0,\beta_1)| \le g^{-\frac{1}{2}}M, \\
&|\mathrm{E}\left\{l_{i\beta_0\sigma_e^2}(\boldsymbol{\omega})-l_{i\beta_0\sigma_e^2}(\boldsymbol{\omega}_0)\right\}| = |m^{-1}\zeta h_i(\beta_0,\beta_1)| \le g^{-\frac{1}{2}}M, \\
&|\mathrm{E}\{l_{i\beta_1\beta_0}(\boldsymbol{\omega})-l_{i\beta_1\beta_0}(\boldsymbol{\omega}_0)\}| = |(\zeta-\dot\zeta)\bar{x}_{i.}| \le g^{-\frac{1}{2}}M, \\
&|\mathrm{E}\{l_{i\beta_1\beta_1}(\boldsymbol{\omega})-l_{i\beta_1\beta_1}(\boldsymbol{\omega}_0)\}| = |(\zeta-\dot\zeta)\bar{x}_{i.}^2| \le g^{-\frac{1}{2}}M, \\
&|\mathrm{E}\left\{l_{i\beta_1\sigma_\alpha^2}(\boldsymbol{\omega})-\mathrm{E}\,l_{i\beta_1\sigma_\alpha^2}(\boldsymbol{\omega}_0)\right\}| = |\zeta^2\bar{x}_{i.}h_i(\beta_0,\beta_1) \le g^{-\frac{1}{2}}M, \\
&|\mathrm{E}\left\{l_{i\beta_1\sigma_e^2}(\boldsymbol{\omega})-\mathrm{E}\,l_{i\beta_1\sigma_e^2}(\boldsymbol{\omega}_0)\right\}| = |m^{-1}\zeta^2 h_i(\beta_0,\beta_1)| \le g^{-\frac{1}{2}}M, \\
&|\mathrm{E}\left\{l_{i\sigma_\alpha^2\beta_0}(\boldsymbol{\omega})-l_{i\sigma_\alpha^2\beta_0}(\boldsymbol{\omega}_0)\right\}| = |\zeta^2 h_i(\beta_0,\beta_1)| \le g^{-\frac{1}{2}}M,
\end{aligned}
$$

$$|\,\mathrm{E}\left\{l_{i\sigma_\alpha^2\beta_1}(\boldsymbol{\omega}) - l_{i\sigma_\alpha^2\beta_1}(\boldsymbol{\omega}_0)\right\}| = |\zeta^2 \bar{x}_{i.} h_i(\beta_0,\beta_1)| \le g^{-\frac{1}{2}} M,$$

$$\begin{aligned}
|\,\mathrm{E}\left\{l_{i\sigma_\alpha^2\sigma_\alpha^2}(\boldsymbol{\omega}) - l_{i\sigma_\alpha^2\sigma_\alpha^2}(\boldsymbol{\omega}_0)\right\}| &= |\frac{1}{2}\zeta^2 - \zeta^3\left\{h_i^2(\beta_0,\beta_1) + \frac{1}{\dot{\zeta}}\right\} + \frac{1}{2}\dot{\zeta}^2| \\
&\le |\frac{1}{2}\zeta^2 - \frac{\zeta^3}{\dot{\zeta}} + \frac{1}{2}\dot{\zeta}^2| + |\zeta^3 h_i^2(\beta_0,\beta_1)| \\
&= |(\zeta^2 - \frac{1}{2}\zeta^2) - \frac{\zeta^3}{\dot{\zeta}} + \frac{1}{2}\dot{\zeta}^2| + |\zeta^3 h_i^2(\beta_0,\beta_1)| \\
&= |\zeta^2(1 - \frac{\zeta}{\dot{\zeta}}) + \frac{1}{2}(\dot{\zeta}+\zeta)(\dot{\zeta}-\zeta)| + |\zeta^3 h_i^2(\beta_0,\beta_1)| \\
&\le |\zeta^2(1 - \frac{\zeta}{\dot{\zeta}})| + \frac{1}{2}|(\dot{\zeta}+\zeta)||(\dot{\zeta}-\zeta)| + |\zeta^3 h_i^2(\beta_0,\beta_1)| \\
&\le g^{-\frac{1}{2}} M
\end{aligned}$$

and

$$|\,\mathrm{E}\left\{l_{i\sigma_\alpha^2\sigma_e^2}(\boldsymbol{\omega}) - l_{i\sigma_\alpha^2\sigma_e^2}(\boldsymbol{\omega}_0)\right\}| = m^{-1}|\frac{1}{2}\zeta^2 - \zeta^3\left\{h_i^2(\beta_0,\beta_1) + \frac{1}{\dot{\zeta}}\right\} + \frac{1}{2}\dot{\zeta}^2| \le g^{-\frac{1}{2}} m^{-1} M.$$

Thus,

$$\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} g^{-\frac{1}{2}} \sum_{i=1}^{g} |\left[\mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega})\right\} - \mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega}_0)\right\}\right](\boldsymbol{\omega}-\boldsymbol{\omega}_0)| = O(g^{-\frac{1}{2}}).$$

The second equation of lemma 6 again requires more care. From (2.2.27), (2.2.28), (2.2.29) and (2.2.30), we have

$$\begin{aligned}
&\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} g^{-\frac{1}{2}} m^{-\frac{1}{2}} \sum_{i=1}^{g}\sum_{j=1}^{m} |\left[\mathrm{E}\left\{\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega})\right\} - \mathrm{E}\left\{\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega}_0)\right\}\right](\boldsymbol{\omega}-\boldsymbol{\omega}_0)| \\
&\le \sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} g^{-\frac{1}{2}} m^{-\frac{1}{2}} \sum_{i=1}^{g}\sum_{j=1}^{m} \left[|\left\{\mathrm{E}\,\frac{l_{ij\beta_2}(\boldsymbol{\omega})}{\partial\boldsymbol{\omega}} - \mathrm{E}\,\frac{l_{ij\beta_2}(\boldsymbol{\omega}_0)}{\partial\boldsymbol{\omega}}\right\}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\right. \\
&\quad \left. + |\left\{\mathrm{E}\,\frac{l_{ij\sigma_e^2}(\boldsymbol{\omega})}{\partial\boldsymbol{\omega}} - \mathrm{E}\,\frac{l_{ij\sigma_e^2}(\boldsymbol{\omega}_0)}{\partial\boldsymbol{\omega}}\right\}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\right] \\
&\le \sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} g^{-\frac{1}{2}} m^{-\frac{1}{2}} \sum_{i=1}^{g}\sum_{j=1}^{m} \left[|\left\{\mathrm{E}\,l_{ij\beta_2\beta_2}(\boldsymbol{\omega}) - \mathrm{E}\,l_{ij\beta_2\beta_2}(\boldsymbol{\omega}_0)\right\}(\beta_2-\dot{\beta}_2)| + |\,\mathrm{E}\,l_{ij\beta_2\sigma_e^2}(\boldsymbol{\omega})(\sigma_e^2 - \dot{\sigma_e}^2)|\right. \\
&\quad + |\,\mathrm{E}\,l_{ij\sigma_e^2\beta_0}(\boldsymbol{\omega})(\beta_0-\dot{\beta}_0)| + |\,\mathrm{E}\,l_{ij\sigma_e^2\beta_1}(\boldsymbol{\omega})(\beta_1-\dot{\beta}_1)| + |\,\mathrm{E}\,l_{ij\sigma_e^2\sigma_\alpha^2}(\boldsymbol{\omega})(\sigma_\alpha^2 - \dot{\sigma_\alpha}^2)| \\
&\quad \left. + |\,\mathrm{E}\,l_{ij\sigma_e^2\beta_2}(\boldsymbol{\omega})(\beta_2-\dot{\beta}_2)| + |\left\{\mathrm{E}\,l_{ij\sigma_e^2\sigma_e^2}(\boldsymbol{\omega}) - El_{ij\sigma_e^2\sigma_e^2}(\boldsymbol{\omega}_0)\right\}(\sigma_e^2 - \dot{\sigma_e}^2)|\right].
\end{aligned}$$

Note that some terms such as $\mathrm{E}\,l_{ij\beta_1\sigma_e^2}(\boldsymbol{\omega}_0) = 0$ have already been set equal to zero in above upper bound. Since $|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)| \le M$, as $g, m \to \infty$, then we have $|\beta_0 - \dot{\beta}_0| \le g^{-\frac{1}{2}}M$, $|\beta_1 - \dot{\beta}_1| \le g^{-\frac{1}{2}}M$, $|\sigma_\alpha^2 - \dot{\sigma_\alpha}^2| \le g^{-\frac{1}{2}}M$, $|\beta_2 - \dot{\beta}_2| \le g^{-\frac{1}{2}}m^{-\frac{1}{2}}M$ and $|\sigma_e^2 - \dot{\sigma_e}^2| \le g^{-\frac{1}{2}}m^{-\frac{1}{2}}M$. Combining (2.2.27), (2.2.28), (2.2.29) and (2.2.30), we obtain

$$
\begin{aligned}
&|\{\mathrm{E}\, l_{ij\beta_2\beta_2}(\boldsymbol{\omega}) - \mathrm{E}\, l_{ij\beta_2\beta_2}(\boldsymbol{\omega}_0)\}(\beta_2 - \dot{\beta}_2)| = |\frac{\sigma_e^2 - \dot{\sigma_e}^2}{\sigma_e^2\dot{\sigma_e}^2} S_{ijw}^x(\beta_2 - \dot{\beta}_2)| \le g^{-1}m^{-1}M,\\
&|\mathrm{E}\, l_{ij\beta_2\sigma_e^2}(\boldsymbol{\omega})(\sigma_e^2 - \dot{\sigma_e}^2)| = |-\frac{1}{\sigma_e^4}(\dot{\beta}_2 - \beta_2)(\sigma_e^2 - \dot{\sigma_e}^2)| \le g^{-1}m^{-1}M,\\
&|\mathrm{E}\, l_{ij\sigma_e^2\beta_0}(\boldsymbol{\omega})(\beta_0 - \dot{\beta}_0)| = |\frac{-1}{(\sigma_e^2 + m\sigma_\alpha^2)^2}[(\dot{\beta}_0 - \beta_0) + \bar{x}_{i.}(\dot{\beta}_1 - \beta_1)](\dot{\beta}_0 - \beta_0)| \le g^{-1}m^{-2}M,\\
&|\mathrm{E}\, l_{ij\sigma_e^2\beta_1}(\boldsymbol{\omega})(\beta_1 - \dot{\beta}_1)| = |\frac{-\bar{x}_{i.}}{(\sigma_e^2 + m\sigma_\alpha^2)^2}[(\dot{\beta}_0 - \beta_0) + \bar{x}_{i.}(\dot{\beta}_1 - \beta_1)](\dot{\beta}_1 - \beta_1)| \le g^{-1}m^{-2}M,\\
&|\mathrm{E}\, l_{ij\sigma_e^2\beta_2}(\boldsymbol{\omega})(\beta_2 - \dot{\beta}_2)| = |\frac{1}{\sigma_e^4} S_{ijw}^x(\dot{\beta}_2 - \beta_2)^2| \le g^{-1}m^{-1}M,
\end{aligned}
$$

$$
\begin{aligned}
&|\mathrm{E}\left\{l_{ij\sigma_e^2\sigma_e^2}(\boldsymbol{\omega}) - \mathrm{E}\, l_{ij\sigma_e^2\sigma_e^2}(\boldsymbol{\omega}_0)\right\}(\sigma_e^2 - \dot{\sigma_e}^2)|\\
&= |\left[\frac{1}{2}m^{-3}\zeta^2 - m^{-3}\zeta^3\left\{h_i^2(\beta_0,\beta_1) + \frac{1}{\dot{\zeta}}\right\} + \frac{m-1}{2m\sigma_e^4} - \frac{1}{\sigma_e^6}\left\{S_{ijw}^x(\dot{\beta}_2 - \beta_2)^2 - \frac{m-1}{m}\dot{\sigma_e}^2\right\}\right.\\
&\qquad \left. + \frac{1}{2}m^{-3}\dot{\zeta}^2 + \frac{m-1}{2m\dot{\sigma_e}^4}\right](\sigma_e^2 - \dot{\sigma_e}^2)|\\
&\le |m^{-3}\left[\frac{1}{2}\zeta^2 - m^{-3}\zeta^3\left\{h_i^2(\beta_0,\beta_1) + \frac{1}{\dot{\zeta}}\right\} + \frac{1}{2}m^{-3}\dot{\zeta}^2\right](\sigma_e^2 - \dot{\sigma_e}^2)|\\
&\quad + |\left[\frac{m-1}{2m\sigma_e^4} - \frac{1}{\sigma_e^6}\left\{S_{ijw}^x(\dot{\beta}_2 - \beta_2)^2 - \frac{m-1}{m}\dot{\sigma_e}^2\right\} + \frac{m-1}{2m\dot{\sigma_e}^4}\right](\sigma_e^2 - \dot{\sigma_e}^2)|\\
&\le g^{-1}m^{-\frac{7}{2}}M + |\left[\frac{m-1}{2m\sigma_e^4} - \frac{1}{\sigma_e^6}\left\{S_{ijw}^x(\dot{\beta}_2 - \beta_2)^2 - \frac{m-1}{m}\dot{\sigma_e}^2\right\} + \frac{m-1}{2m\dot{\sigma_e}^4}\right](\sigma_e^2 - \dot{\sigma_e}^2)|\\
&\le g^{-1}m^{-\frac{7}{2}}M + |\left\{\frac{m-1}{2m\sigma_e^4} - \frac{1}{\sigma_e^6}\frac{m-1}{m}\dot{\sigma_e}^2 + \frac{m-1}{2m\dot{\sigma_e}^4}\right\}(\sigma_e^2 - \dot{\sigma_e}^2)| + |\frac{1}{\sigma_e^6}S_{ijw}^x(\dot{\beta}_2 - \beta_2)^2(\sigma_e^2 - \dot{\sigma_e}^2)|\\
&\le g^{-1}m^{-\frac{7}{2}}M + g^{-\frac{3}{2}}m^{-\frac{3}{2}}M + |\left\{\frac{m-1}{2m\sigma_e^4} - \frac{1}{\sigma_e^6}\frac{m-1}{m}\dot{\sigma_e}^2 + \frac{m-1}{2m\dot{\sigma_e}^4}\right\}(\sigma_e^2 - \dot{\sigma_e}^2)|\\
&= g^{-1}m^{-\frac{7}{2}}M + g^{-\frac{3}{2}}m^{-\frac{3}{2}}M + \frac{m-1}{m}|\left\{\frac{1}{\sigma_e^4} - \frac{1}{2\sigma_e^4} - \frac{1}{\sigma_e^6}\dot{\sigma_e}^2 + \frac{1}{2\dot{\sigma_e}^4}\right\}(\sigma_e^2 - \dot{\sigma_e}^2)|\\
&\le g^{-1}m^{-\frac{7}{2}}M + g^{-\frac{3}{2}}m^{-\frac{3}{2}}M + \frac{m-1}{m}|\left\{\frac{1}{\sigma_e^4} - \frac{1}{2\sigma_e^4} - \frac{1}{\sigma_e^6}\dot{\sigma_e}^2 + \frac{1}{2\dot{\sigma_e}^4}\right\}(\sigma_e^2 - \dot{\sigma_e}^2)|\\
&\le g^{-1}m^{-\frac{7}{2}}M + g^{-\frac{3}{2}}m^{-\frac{3}{2}}M + \frac{m-1}{m}|\frac{1}{\sigma_e^4}(1 - \frac{\dot{\sigma_e}^2}{\sigma_e^2}) + \frac{1}{2}(\frac{1}{\dot{\sigma_e}^4} - \frac{1}{\sigma_e^4})||(\sigma_e^2 - \dot{\sigma_e}^2)|\\
&\le g^{-1}m^{-\frac{7}{2}}M + g^{-\frac{3}{2}}m^{-\frac{3}{2}}M + g^{-1}m^{-1}M\\
&\le g^{-1}m^{-1}M,
\end{aligned}
$$

and

$$\begin{aligned}
&|\left\{\mathrm{E}\, l_{ij\sigma_e^2\sigma_\alpha^2}(\boldsymbol{\omega}) - \mathrm{E}\, l_{ij\sigma_e^2\sigma_\alpha^2}(\boldsymbol{\omega}_0)\right\}(\sigma_\alpha^2 - \dot{\sigma_\alpha}^2)| \\
&= |[\frac{1}{2}m^{-2}\zeta^2 - m^{-2}\zeta^3\left\{h_i^2(\beta_0,\beta_1) + \frac{1}{\dot{\zeta}}\right\} + \frac{1}{2}m^{-2}\dot{\zeta}^2](\sigma_\alpha^2 - \dot{\sigma_\alpha}^2)| \\
&\le m^{-2}|\frac{1}{2}\zeta^2 - \zeta^3\left\{h_i^2(\beta_0,\beta_1) + \frac{1}{\dot{\zeta}}\right\} + \frac{1}{2}\dot{\zeta}^2||(\sigma_\alpha^2 - \dot{\sigma_\alpha}^2)|| \\
&\le g^{-1}m^{-2}M.
\end{aligned}$$

Thus,

$$\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} g^{-\frac{1}{2}}m^{-\frac{1}{2}}\sum_{i=1}^{g}\sum_{j=1}^{m}|\left[\mathrm{E}\left\{\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega})\right\} - \mathrm{E}\,\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega}_0)\right](\boldsymbol{\omega}-\boldsymbol{\omega}_0)| = O(g^{-\frac{1}{2}}m^{-\frac{1}{2}}).$$

□

Lemma 7. *Suppose Condition A holds. Then, as $g, m \to \infty$,*

$$\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} |\mathbf{K}^{-\frac{1}{2}}\left[\psi(\boldsymbol{\omega}) - \psi(\boldsymbol{\omega}_0) - \mathrm{E}\left\{\psi(\boldsymbol{\omega}) - \psi(\boldsymbol{\omega}_0)\right\}\right]| = o_p(1). \tag{2.3.3}$$

Proof. Decompose the cube $\mathcal{B} = \{\boldsymbol{\omega} : |\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)| \le M\}$ into smaller cubes $\mathcal{C} = \{\mathcal{C}(t_k)\}$, where M is a constant and $\{\mathcal{C}(t_k)\}$ is a cube containing $\boldsymbol{\omega} \in \mathcal{C}(t_k) = \{\boldsymbol{\omega} : |\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)| \le Mg^{-\frac{1}{4}}\}$. Let $N = g^{\frac{1}{4}}$.We first show that lemma 7 holds over the $\mathcal{C}$ of fixed $t_k = (t_{k1}, t_{k2}, t_{k3}, t_{k4}, t_{k5})^T$. Using Chebyshev's inequality, for any $\eta > 0$, we have

$$\begin{aligned}
&\sum_{k=1}^{N}\Pr\left(|\mathbf{K}^{-\frac{1}{2}}\left[\psi(t_k) - \psi(\boldsymbol{\omega}_0) - \mathrm{E}\left\{\psi(t_k) - \psi(\boldsymbol{\omega}_0)\right\}\right]| > \eta\right) \\
&\le \eta^{-2}\sum_{k=1}^{N}\mathrm{E}\,|\mathbf{K}^{-\frac{1}{2}}\left[\psi(t_k) - \psi(\boldsymbol{\omega}_0) - \mathrm{E}\left\{\psi(t_k) - \psi(\boldsymbol{\omega}_0)\right\}\right]|^2 \\
&= \eta^{-2}\sum_{k=1}^{N}\mathrm{E}\,|g^{-\frac{1}{2}}\left[\psi^{(b)}(t_k) - \psi^{(b)}(\boldsymbol{\omega}_0) - \mathrm{E}\left\{\psi^{(b)}(t_k) - \psi^{(b)}(\boldsymbol{\omega}_0)\right\}\right]|^2 \\
&\quad + \eta^{-2}\sum_{k=1}^{N}\mathrm{E}\,|g^{-\frac{1}{2}}m^{-\frac{1}{2}}\left[\psi^{(w)}(t_k) - \psi^{(w)}(\boldsymbol{\omega}_0) - \mathrm{E}\left\{\psi^{(w)}(t_k) - \psi^{(w)}(\boldsymbol{\omega}_0)\right\}\right]|^2 \qquad (2.3.4) \\
&= \eta^{-2}g^{-1}\sum_{k=1}^{N}\mathrm{E}\left[|\psi^{(b)}(t_k) - \psi^{(b)}(\boldsymbol{\omega}_0) - \mathrm{E}\left\{\psi^{(b)}(t_k) - \psi^{(b)}(\boldsymbol{\omega}_0)\right\}|^2\right] \\
&\quad + \eta^{-2}g^{-1}m^{-1}\sum_{k=1}^{N}\mathrm{E}\left[|\psi^{(w)}(t_k) - \psi^{(w)}(\boldsymbol{\omega}_0) - \mathrm{E}\left\{\psi^{(w)}(t_k) - \psi^{(w)}(\boldsymbol{\omega}_0)\right\}|^2\right].
\end{aligned}$$

For the first term of (2.3.4), using Lemma 2, Lemma 4 and a Taylor expansion, we have

$$
\begin{aligned}
&\eta^{-2}g^{-1}\sum_{k=1}^{N}\mathrm{E}\left[|\psi^{(b)}(t_k)-\psi^{(b)}(\boldsymbol{\omega}_0)-\mathrm{E}\left\{\psi^{(b)}(t_k)-\psi^{(b)}(\boldsymbol{\omega}_0)\right\}|^2\right]\\
&\leq \eta^{-2}g^{-1}\sum_{k=1}^{N}\mathrm{E}\left\{|\psi^{(b)}(t_k)-\psi^{(b)}(\boldsymbol{\omega}_0)|^2\right\}\\
&= \eta^{-2}g^{-1}\sum_{k=1}^{N}\mathrm{E}\left[|\sum_{i=1}^{g}\left\{\psi_i^{(b)}(t_k)-\psi_i^{(b)}(\boldsymbol{\omega}_0)\right\}|^2\right]\\
&\leq \eta^{-2}g^{-1}\sum_{k=1}^{N}\sum_{i=1}^{g}\mathrm{E}\left\{|\psi_i^{(b)}(t_k)-\psi_i^{(b)}(\boldsymbol{\omega}_0)|^2\right\}\\
&= \eta^{-2}g^{-1}\sum_{k=1}^{N}\sum_{i=1}^{g}\mathrm{E}\left\{|\nabla\psi_i^{(b)}(\tilde{\boldsymbol{\omega}})(t_k-\boldsymbol{\omega}_0)|^2\right\},
\end{aligned}
$$

where $\tilde{\boldsymbol{\omega}}$ is between t_k and $\boldsymbol{\omega}_0$. Because $\|\mathbf{K}^{-\frac{1}{2}}\| \leq g^{-\frac{1}{2}}L$ for some constant $L > 0$, it follows that

$$
\begin{aligned}
&\eta^{-2}g^{-1}\sum_{k=1}^{N}\sum_{i=1}^{g}\mathrm{E}\left\{|\nabla\psi_i^{(b)}(\tilde{\boldsymbol{\omega}})(t_k-\boldsymbol{\omega}_0)|^2\right\}\\
&\leq \eta^{-2}g^{-1}\sum_{k=1}^{N}\sum_{i=1}^{g}\mathrm{E}\left\{\|\nabla\psi_i^{(b)}(\tilde{\boldsymbol{\omega}})\|^2\|\mathbf{K}^{-\frac{1}{2}}\|^2|\mathbf{K}^{\frac{1}{2}}(t_k-\boldsymbol{\omega}_0)|^2\right\}\\
&\leq \eta^{-2}g^{-2}M^2L^2\sum_{k=1}^{N}\sum_{i=1}^{g}\mathrm{E}\,\|\nabla\psi_i^{(b)}(\tilde{\boldsymbol{\omega}})\|^2\\
&= O_p(g^{-\frac{3}{4}})\\
&= o_p(1).
\end{aligned}
$$

Similarly, for the second term of (2.3.4), using Lemma 2, Lemma 4 and applying a Taylor expansion, we obtain

$$
\begin{aligned}
&\eta^{-2}g^{-1}m^{-1}\sum_{k=1}^{N}\mathrm{E}\left[|\psi^{(w)}(t_k)-\psi^{(w)}(\boldsymbol{\omega}_0)-\mathrm{E}\left\{\psi^{(w)}(t_k)-\psi^{(w)}(\boldsymbol{\omega}_0)\right\}|^2\right]\\
&\leq \eta^{-2}g^{-1}m^{-1}\sum_{k=1}^{N}\mathrm{E}\left\{|\psi^{(w)}(t_k)-\psi^{(w)}(\boldsymbol{\omega}_0)|^2\right\}\\
&= \eta^{-2}g^{-1}m^{-1}\sum_{k=1}^{N}\mathrm{E}\left[|\sum_{i=1}^{g}\sum_{j=1}^{m}\left\{\psi_{ij}^{(w)}(t_k)-\psi_{ij}^{(w)}(\boldsymbol{\omega}_0)\right\}|^2\right]\\
&\leq \eta^{-2}g^{-1}m^{-1}\sum_{k=1}^{N}\sum_{i=1}^{g}\sum_{j=1}^{m}\mathrm{E}\left\{|\psi_{ij}^{(w)}(t_k)-\psi_{ij}^{(w)}(\boldsymbol{\omega}_0)|^2\right\}\\
&= \eta^{-2}g^{-1}m^{-1}\sum_{k=1}^{N}\sum_{i=1}^{g}\sum_{j=1}^{m}\mathrm{E}\left\{|\nabla\psi_{ij}^{(w)}(\tilde{\boldsymbol{\omega}})(t_k-\boldsymbol{\omega}_0)|^2\right\}\\
&\leq \eta^{-2}g^{-1}m^{-1}\sum_{k=1}^{N}\sum_{i=1}^{g}\sum_{j=1}^{m}\mathrm{E}\left\{\|\nabla\psi_{ij}^{(w)}(\tilde{\boldsymbol{\omega}})\|^2\|\mathbf{K}^{-\frac{1}{2}}\|^2|\mathbf{K}^{\frac{1}{2}}(t_k-\boldsymbol{\omega}_0)|^2\right\}\\
&\leq \eta^{-2}g^{-2}m^{-1}M^2L^2\sum_{k=1}^{N}\sum_{i=1}^{g}\sum_{j=1}^{m}\mathrm{E}\left\{\|\nabla\psi_{ij}^{(w)}(\tilde{\boldsymbol{\omega}})\|^2\right\}\\
&= O_p(g^{-\frac{3}{4}})\\
&= o_p(1).
\end{aligned}
$$

Next, we show that the difference between taking the supremum over a fine grid of points and over $\mathcal{B}$ is small. By using a Taylor expansion, we get

$$
\begin{aligned}
&\max_{1\leq k\leq N}\sup_{\boldsymbol{\omega}\in\mathcal{C}(t_k)}|\mathbf{K}^{-\frac{1}{2}}\left[\psi(\boldsymbol{\omega})-\psi(t_k)-\mathrm{E}\left\{\psi(\boldsymbol{\omega})-\psi(t_k)\right\}\right]|\\
&\leq \max_{1\leq k\leq N}\sup_{\boldsymbol{\omega}\in\mathcal{C}(t_k)}g^{-\frac{1}{2}}|\sum_{i=1}^{g}\left[\psi_i^{(b)}(\boldsymbol{\omega})-\psi_i^{(b)}(t_k)-\mathrm{E}\left\{\psi_i^{(b)}(\boldsymbol{\omega})-\psi_i^{(b)}(t_k)\right\}\right]|\\
&\quad+\max_{1\leq k\leq N}\sup_{\boldsymbol{\omega}\in\mathcal{C}(t_k)}g^{-\frac{1}{2}}m^{-\frac{1}{2}}|\sum_{i=1}^{g}\sum_{j=1}^{m}\left[\psi_{ij}^{(w)}(\boldsymbol{\omega})-\psi_{ij}^{(w)}(t_k)-\mathrm{E}\left\{\psi_{ij}^{(w)}(\boldsymbol{\omega})-\psi_{ij}^{(w)}(t_k)\right\}\right]|\\
&\leq \max_{1\leq k\leq N}\sup_{\boldsymbol{\omega}\in\mathcal{C}(t_k)}g^{-\frac{1}{2}}\sum_{i=1}^{g}|\psi_i^{(b)}(\boldsymbol{\omega})-\psi_i^{(b)}(t_k)-\mathrm{E}\left\{\psi_i^{(b)}(\boldsymbol{\omega})-\psi_i^{(b)}(t_k)\right\}|\\
&\quad+\max_{1\leq k\leq N}\sup_{\boldsymbol{\omega}\in\mathcal{C}(t_k)}g^{-\frac{1}{2}}m^{-\frac{1}{2}}\sum_{i=1}^{g}\sum_{j=1}^{m}|\psi_{ij}^{(w)}(\boldsymbol{\omega})-\psi_{ij}^{(w)}(t_k)-\mathrm{E}\left\{\psi_{ij}^{(w)}(\boldsymbol{\omega})-\psi_{ij}^{(w)}(t_k)\right\}|\\
&\leq \max_{1\leq k\leq N}\sup_{\boldsymbol{\omega}\in\mathcal{C}(t_k)}g^{-\frac{1}{2}}\sum_{i=1}^{g}|\left[\nabla\psi_i^{(b)}(\tilde{\boldsymbol{\omega}}_1)-\mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega}_1^*)\right\}\right](\boldsymbol{\omega}-t_k)|\\
&\quad+\max_{1\leq k\leq N}\sup_{\boldsymbol{\omega}\in\mathcal{C}(t_k)}g^{-\frac{1}{2}}m^{-\frac{1}{2}}\sum_{i=1}^{g}\sum_{j=1}^{m}|\left[\nabla\psi_{ij}^{(w)}(\tilde{\boldsymbol{\omega}}_2)-\mathrm{E}\left\{\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega}_2^*)\right\}\right](\boldsymbol{\omega}-t_k)|
\end{aligned}
\tag{2.3.5}
$$

where $\tilde{\boldsymbol{\omega}}_1, \tilde{\boldsymbol{\omega}}_2, \boldsymbol{\omega}_1^*$ and $\boldsymbol{\omega}_2^*$ are between t_k and $\boldsymbol{\omega}$. The result follows from lemma 5. □

Lemma 8. *Suppose Condition A holds. Then, as $g, m \to \infty$,*

$$\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} |\mathbf{K}^{-\frac{1}{2}}\left[\mathrm{E}\left\{\psi(\boldsymbol{\omega})-\psi(\boldsymbol{\omega}_0)\right\}-\mathrm{E}\left\{\nabla\psi(\boldsymbol{\omega}_0)\right\}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)\right]| = o_p(1). \tag{2.3.6}$$

Proof. Write

$$\begin{aligned}
&\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} |\mathbf{K}^{-\frac{1}{2}}\left[\mathrm{E}\left\{\psi(\boldsymbol{\omega})-\psi(\boldsymbol{\omega}_0)\right\}-\mathrm{E}\left\{\nabla\psi(\boldsymbol{\omega}_0)\right\}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)\right]| \\
\le& \sup_{|\mathbf{K}^{-\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} g^{-\frac{1}{2}}|\sum_{i=1}^{g}\left[\mathrm{E}\left\{\psi_i^{(b)}(\boldsymbol{\omega})-\psi_i^{(b)}(\boldsymbol{\omega}_0)\right\}-\mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega}_0)\right\}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)\right]| \\
&+\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} g^{-\frac{1}{2}}m^{-\frac{1}{2}}|\sum_{i=1}^{g}\sum_{j=1}^{m}\left[\mathrm{E}\left\{\psi_{ij}^{(w)}(\boldsymbol{\omega})-\psi_{ij}^{(w)}(\boldsymbol{\omega}_0)\right\}-\mathrm{E}\left\{\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega}_0)\right\}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)\right]| \\
\le& \sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} g^{-\frac{1}{2}}\sum_{i=1}^{g}|\,\mathrm{E}\left\{\psi_i^{(b)}(\boldsymbol{\omega})-\psi_i^{(b)}(\boldsymbol{\omega}_0)\right\}-\mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega}_0)\right\}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)| \\
&+\sup_{|(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} g^{-\frac{1}{2}}m^{-\frac{1}{2}}\sum_{i=1}^{g}\sum_{j=1}^{m}|\,\mathrm{E}\left\{\psi_{ij}^{(w)}(\boldsymbol{\omega})-\psi_{ij}^{(w)}(\boldsymbol{\omega}_0)\right\}-\mathrm{E}\left\{\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega}_0)\right\}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)| \\
=& \sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} g^{-\frac{1}{2}}\sum_{i=1}^{g}|\left[\mathrm{E}\left\{\nabla\psi_i^{(b)}(\tilde{\boldsymbol{\omega}}_1)\right\}-\mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega}_0)\right\}\right](\boldsymbol{\omega}-\boldsymbol{\omega}_0)| \\
&+\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} g^{-\frac{1}{2}}m^{-\frac{1}{2}}\sum_{i=1}^{g}\sum_{j=1}^{m}|\left[\mathrm{E}\left\{\nabla\psi_{ij}^{(w)}(\tilde{\boldsymbol{\omega}}_2)\right\}-\mathrm{E}\left\{\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega}_0)\right\}\right](\boldsymbol{\omega}-\boldsymbol{\omega}_0)|,
\end{aligned}$$

where $\tilde{\boldsymbol{\omega}}_1$ and $\tilde{\boldsymbol{\omega}}_2$ are between $\boldsymbol{\omega}$ and $\boldsymbol{\omega}_0$. The result follows from lemma 6 . □

2.3.1 Proof of the main theorem

Proof. We write

$$\begin{aligned}
\mathbf{K}^{-\frac{1}{2}}\psi(\boldsymbol{\omega}) =& \mathbf{K}^{-\frac{1}{2}}\left[\psi(\boldsymbol{\omega})-\psi(\boldsymbol{\omega}_0)-\mathrm{E}\left\{\psi(\boldsymbol{\omega})-\psi(\boldsymbol{\omega}_0)\right\}\right]+\mathbf{K}^{-\frac{1}{2}}\left\{\psi(\boldsymbol{\omega}_0)-\xi(\boldsymbol{\omega}_0)\right\} \\
&+\mathbf{K}^{-\frac{1}{2}}\left[\mathrm{E}\left\{\psi(\boldsymbol{\omega})-\psi(\boldsymbol{\omega}_0)-\nabla\psi(\boldsymbol{\omega}_0)(\boldsymbol{\omega}-\boldsymbol{\omega}_0)\right\}\right] \\
&+\mathbf{K}^{-\frac{1}{2}}[\xi(\boldsymbol{\omega}_0)+\mathrm{E}\{\nabla\psi(\boldsymbol{\omega}_0)\}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)].
\end{aligned}$$

From (2.2.9), we have

$$\mathbf{K}^{-\frac{1}{2}}\left\{\psi(\boldsymbol{\omega}_0)-\xi(\boldsymbol{\omega}_0)\right\} = o_p(1).$$

From lemma 7 and lemma 8, we have

$$\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} |\mathbf{K}^{-\frac{1}{2}}\left[\psi(\boldsymbol{\omega})-\psi(\boldsymbol{\omega}_0)-\mathrm{E}\left\{\psi(\boldsymbol{\omega})-\psi(\boldsymbol{\omega}_0)\right\}\right]| = o_p(1) \qquad \text{and}$$

$$\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} |\mathbf{K}^{-\frac{1}{2}}\left[\mathrm{E}\left\{\psi(\boldsymbol{\omega})-\psi(\boldsymbol{\omega}_0)\right\}-\mathrm{E}\left\{\nabla\psi(\boldsymbol{\omega}_0)\right\}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)\right]| = o_p(1).$$

Thus, combining above equations, we can see that uniformly on $\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega} - \boldsymbol{\omega}_0) \leq M$,

$$\mathbf{K}^{-\frac{1}{2}}\psi(\boldsymbol{\omega}) = \mathbf{K}^{-\frac{1}{2}}\xi(\boldsymbol{\omega}_0)\mathbf{B}_n\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega} - \boldsymbol{\omega}_0) + o_p(1), \tag{2.3.7}$$

where $\mathbf{B}_n = -\mathbf{K}^{-\frac{1}{2}}\,\mathrm{E}\left\{\nabla\psi(\boldsymbol{\omega}_0)\right\}\mathbf{K}^{-\frac{1}{2}}$. The right-hand side of (2.3.7) is mainly determined by the negative of the quadratic form of $\mathbf{B}_n$ and is negative for M sufficiently large. Therefore, according to Result 6.3.4 of Ortega and Rheinboldt (1973), a solution to the estimating equations exists in probability and satisfies $|\mathbf{K}^{-\frac{1}{2}}(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega}_0)| = O_p(1)$. The asymptotic representation follows from lemma 7 and lemma 8. Lemma 3 implies that the Lyapounov condition holds and hence the asymptotic normality follows. □

2.4 Restricted Maximum Likelihood Estimating Equations

In order to better introduce REML asymptotic properties, we denote $\mathbf{X} = [\mathbf{1}_{gm}, \bar{\mathbf{x}}, \mathbf{x} - \bar{\mathbf{x}}]$, where $\bar{\mathbf{x}}^T = [\mathbf{1}_m^T\bar{x}_1, \ldots, \mathbf{1}_m^T\bar{x}_g]$ and $\mathbf{x}^T = [\mathbf{x}_1^T, \ldots, \mathbf{x}_g^T]$, then we rewrite (2.1.3) in general matrix form as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\boldsymbol{\alpha} + \mathbf{Z}_2\mathbf{e}. \tag{2.4.1}$$

In this case, $\mathbf{Z}_1 = \mathbf{I}_g \otimes \mathbf{1}_m$ and $\mathbf{Z}_2 = \mathbf{I}_{gm}$. To overcome bias problem, Patterson and Thompson (1971) and Patterson and Thompson (1974) introduced the restricted maximum likelihood (REML) method, which involves maximizing the likelihood of independent contrasts of the data rather than itself. The use of data contrasts results in a modified profile likelihood from which $\boldsymbol{\beta}$ has been eliminated. Rather than using $\mathbf{y}$ directly, REML is based on linear combinations of the elements of $\mathbf{y}$, chosen in such a way that those combinations do not contain any fixed effects, whatever their value. These linear combinations turn out to be equivalent to residuals obtained after fitting the fixed effects. We take the observation $\mathbf{y}$ and perform the transformation of $\mathbf{y}$ to $\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix} = \mathbf{L}^T\mathbf{y}$ where the matrix $\mathbf{L} = [\mathbf{L}_1\mathbf{L}_2]$ consists of two specially chosen sub-matrices: a $n \times r$ matrix $\mathbf{L}_1$ and a $n \times (n-p)$ matrix $\mathbf{L}_2$. The two properties we need for these sub-matrices,(note that there are p fixed effects in the LMM, and so in this case, $p = 3$) are $\mathbf{L}_1^T\mathbf{X} = \mathbf{I}_p$ and $\mathbf{L}_2^T\mathbf{X} = \mathbf{0}$. Under these conditions, Harville (1974) obtained the residual log-likelihood function,

$$l_R(\mathbf{y};\boldsymbol{\theta}) = -\frac{1}{2}(n-r)\log 2\pi - \frac{1}{2}\log\|\mathbf{V}(\boldsymbol{\theta})\| - \frac{1}{2}\log\|\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}\| - \frac{1}{2}\mathbf{y}^T\mathbf{P}(\boldsymbol{\theta})\mathbf{y}, \tag{2.4.2}$$

where $\mathbf{P}(\boldsymbol{\theta}) = \mathbf{V}^{-1}(\boldsymbol{\theta}) - \mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}\{\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}\}^{-1}\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})$.

To maximize $l_R(\mathbf{y};\boldsymbol{\theta})$, we differentiate (2.4.2) with respect to σ_α^2 and σ_e^2, respectively. Specially, we calculate

$$\begin{aligned}
&-\frac{1}{2}\frac{\partial \log \|\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}\|}{\partial\sigma_\alpha^2} \\
&= \frac{1}{2}\operatorname{tr}\left[\left\{\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}\right\}^{-1}\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{Z}_1\mathbf{Z}_1^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}\right] \\
&= \frac{m}{(\sigma_e^2 + m\sigma_\alpha^2)}
\end{aligned} \tag{2.4.3}$$

and

$$\begin{aligned}
&-\frac{1}{2}\frac{\partial \log \|\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}\|}{\partial\sigma_e^2} \\
&= \frac{1}{2}\operatorname{tr}\left[\left\{\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}\right\}^{-1}\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{Z}_2\mathbf{Z}_2^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}\right] \\
&= \frac{1}{(\sigma_e^2 + m\sigma_\alpha^2)} + \frac{1}{2\sigma_e^2}.
\end{aligned} \tag{2.4.4}$$

Furthermore, we can write

$$-\frac{1}{2}\mathbf{y}^T\mathbf{P}(\boldsymbol{\theta})\mathbf{y} = -\frac{1}{2}\left\{\mathbf{y} - \mathbf{X}\boldsymbol{\beta}(\boldsymbol{\theta})\right\}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\left\{\mathbf{y} - \mathbf{X}\boldsymbol{\beta}(\boldsymbol{\theta})\right\},$$

which is maximized over $\boldsymbol{\beta}(\boldsymbol{\theta})$ at

$$\boldsymbol{\beta}(\boldsymbol{\theta}) = \{\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}\}^{-1}\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{y} = \begin{bmatrix}\beta_0(\boldsymbol{\theta})\\ \beta_1(\boldsymbol{\theta})\\ \beta_2(\boldsymbol{\theta})\end{bmatrix} = \begin{bmatrix}\bar{y}_{..} - \frac{S_b^{xy}}{S_b^x}\bar{x}_{..}\\ \frac{S_b^{xy}}{S_b^x}\\ \frac{S_w^{xy}}{S_w^x}\end{bmatrix},$$

such that

$$-\frac{1}{2}\mathbf{y}^T\mathbf{P}(\boldsymbol{\theta})\mathbf{y} = -\frac{1}{2\sigma_e^2}\left\{S_w^y - \frac{(S_w^{xy})^2}{S_w^x}\right\} - \frac{1}{2(\sigma_e^2 + m\sigma_\alpha^2)}\left\{S_b^y - \frac{(S_b^{xy})^2}{S_b^x}\right\}.$$

Thus, we get

$$-\frac{1}{2}\mathbf{y}^T\frac{\partial\mathbf{P}(\boldsymbol{\theta})}{\partial\sigma_\alpha^2}\mathbf{y} = \frac{m}{2(\sigma_e^2 + m\sigma_\alpha^2)^2}\left\{S_b^y - \frac{(S_b^{xy})^2}{S_b^x}\right\} \tag{2.4.5}$$

and

$$-\frac{1}{2}\mathbf{y}^T\frac{\partial\mathbf{P}(\boldsymbol{\theta})}{\partial\sigma_e^2}\mathbf{y} = \frac{1}{2\sigma_e^4}\left\{S_w^x - \frac{(S_w^{xy})^2}{S_w^x}\right\} + \frac{1}{2(\sigma_e^2 + m\sigma_\alpha^2)^2}\left\{S_b^y - \frac{(S_b^{xy})^2}{S_b^x}\right\}. \tag{2.4.6}$$

Combining the results of (2.4.3), (2.4.4), (2.4.5) and (2.4.6), we can write the estimating equations as $\psi_R(\boldsymbol{\theta}) = (l_{R\sigma^2_\alpha}(\boldsymbol{\theta}), l_{R\sigma^2_e}(\boldsymbol{\theta}))^T$, where

$$l_{R\sigma^2_\alpha}(\boldsymbol{\theta}) = -\frac{1}{2}\frac{gm(1-2g^{-1})}{(\sigma^2_e + m\sigma^2_\alpha)} + \frac{m}{2(\sigma^2_e + m\sigma^2_\alpha)^2}\left\{S^y_b - \frac{(S^{xy}_b)^2}{S^x_b}\right\}$$

$$l_{R\sigma^2_e}(\boldsymbol{\theta}) = -\frac{g-2}{2(\sigma^2_e + m\sigma^2_\alpha)} - \frac{(gm-g-1)}{2\sigma^2_e} + \frac{1}{2\sigma^4_e}\left\{S^y_w - \frac{(S^{xy}_w)^2}{S^x_w}\right\} + \frac{1}{2(\sigma^2_e + m\sigma^2_\alpha)^2}\left\{S^y_b - \frac{(S^{xy}_b)^2}{S^x_b}\right\}. \tag{2.4.7}$$

2.4.1 Approximating the REML estimating equations evaluated at $\boldsymbol{\theta}_0$

The final central limit theorem is strongly driven by the behavior of the estimating equation evaluated at $\boldsymbol{\theta}_0$. We therefore start by developing a simple approximation to $\psi_R(\boldsymbol{\theta}_0)$.

First, we obtain approximation for $S^y_w - \frac{(S^{xy}_w)^2}{S^x_w}$. From (2.2.3) and (2.2.5), we get

$$S^y_w - \frac{(S^{xy}_w)^2}{S^x_w} = \sum_{i=1}^{g}\sum_{j=1}^{m}(e^2_{ij} - \dot{\sigma_e}^2) + mg\dot{\sigma_e}^2 - \frac{\{\sum_{i=1}^{g}\sum_{j=1}^{m}(x_{ij}-\bar{x}_{i.})e_{ij}\}^2}{S^x_w} - m\sum_{i=1}^{g}\bar{e}^2_{i.} \tag{2.4.8}$$

and

$$\mathrm{E}\{S^y_w - \frac{(S^{xy}_w)^2}{S^x_w}\} = mg\dot{\sigma_e}^2 - (1+g)\dot{\sigma_e}^2. \tag{2.4.9}$$

Let $u_{ij} = (x_{ij} - \bar{x}_{i.})e_{ij}$ and $\bar{u}_{i.} = \frac{1}{m}\sum_{j=1}^{m} u_{ij}$, so we can rewrite (2.4.8) as

$$\begin{aligned}
&S^y_w - \frac{(S^{xy}_w)^2}{S^x_w} \\
&= \sum_{i=1}^{g}\sum_{j=1}^{m} e^2_{ij} - m\sum_{i=1}^{g}\bar{e}^2_{i.} - \frac{(\sum_{i=1}^{g}\sum_{j=1}^{m}u_{ij})^2}{S^x_w} \\
&= \sum_{i=1}^{g}\sum_{j=1}^{m} e^2_{ij} - m\sum_{i=1}^{g}\bar{e}^2_{i.} - \frac{\sum_{i=1}^{g}\sum_{j=1}^{m}(u^2_{ij} - \sum_{r\neq i}^{g}u_{ij}u_{rj} + \sum_{s\neq j}^{m}u_{ij}u_{is} + \sum_{r\neq i}^{g}\sum_{s\neq j}^{m}u_{ij}u_{rs})}{S^x_w} \\
&= \sum_{i=1}^{g}\sum_{j=1}^{m}\left\{e^2_{ij} - \bar{e}^2_{i.} - \frac{u^2_{ij} - \sum_{r\neq i}^{g}u_{ij}u_{rj} + \sum_{s\neq j}^{m}u_{ij}u_{is} + \sum_{r\neq i}^{g}\sum_{s\neq j}^{m}u_{ij}u_{rs}}{S^x_w}\right\} \\
&= \sum_{i=1}^{g}\sum_{j=1}^{m}\left\{S^x_{ijw} - \frac{(S^{xy}_{ijw})^2 + \sum_{r\neq i}^{g}S^{xy}_{ijw}S^{xy}_{rjw} + \sum_{s\neq j}^{m}S^{xy}_{ijw}S^{xy}_{isw} + \sum_{r\neq i}^{g}\sum_{s\neq j}^{m}S^{xy}_{ijw}S^{xy}_{rsw}}{S^x_w}\right\}.
\end{aligned} \tag{2.4.10}$$

Note that $\mathrm{E}(u_{ij}) = 0$, $\mathrm{E}(u^2_{ij}) = (x_{ij} - \bar{x}_{i.})^2\dot{\sigma_e}^2$ and $\mathrm{E}(u^4_{ij}) = (x_{ij} - \bar{x}_{i.})^4\,\mathrm{E}(e^4_{ij})$. Since Condition A3 implies that $\lim_{g\to\infty}\frac{1}{g}\sum_{i=1}^{g}\bar{x}^k_{i.} \to c_k$ for $k = 1, 2, 3, 4$, then we obtain $\lim_{g,m\to\infty}\sum_{i=1}^{g}\sum_{j=1}^{m}(x_{ij}-$

$\bar{x}_{i.})^4 \operatorname{E}(e_{ij}^4) = O(gm)$. Using Lemma 1, we obtain

$$\begin{aligned}
&\operatorname{Var}\left[\frac{\left\{\sum_{i=1}^{g}\sum_{j=1}^{m}(x_{ij}-\bar{x}_{i.})e_{ij}\right\}^2}{S_w^x}\right] \\
&= \operatorname{Var}\left\{\frac{(\sum_{i=1}^{g}\sum_{j=1}^{m}u_{ij})^2}{S_w^x}\right\} \\
&= \frac{1}{(S_w^x)^2}\left[\operatorname{E}(\sum_{i=1}^{g}\sum_{j=1}^{m}u_{ij})^4 - \left\{\operatorname{E}(\sum_{i=1}^{g}\sum_{j=1}^{m}u_{ij})^2\right\}^2\right] \\
&= \frac{1}{(S_w^x)^2}\sum_{i=1}^{g}\sum_{j=1}^{m}\operatorname{E}u_{ij}^4 + \frac{3}{(S_w^x)^2}\sum_{i=1}^{g}\sum_{j=1}^{m}\sum_{r\neq j}^{m}\operatorname{E}(u_{ij}^2)\operatorname{E}(u_{ir}^2) + \frac{3}{(S_w^x)^2}\sum_{i=1}^{g}\sum_{k\neq i}^{g}\sum_{j=1}^{m}\sum_{s=1}^{m}\operatorname{E}(u_{ij}^2)\operatorname{E}(u_{ks}^2) \\
&\qquad - \frac{1}{(S_w^x)^2}\sum_{i=1}^{g}\sum_{r=1}^{g}\sum_{j=1}^{m}\sum_{s=1}^{m}\operatorname{E}(u_{ij}^2)\operatorname{E}(u_{rs}^2) \\
&= O(g^{-1}m^{-1}) + O(g^{-1}) + O(1) + O(1) \\
&= O(1).
\end{aligned}$$

From the last line of (2.4.10), we can find that

$$S_{ijw}^x - \frac{(S_{ijw}^{xy})^2 + \sum_{r\neq i}^{g} S_{ijw}^{xy}S_{rjw}^{xy} + \sum_{s\neq j}^{m} S_{ijw}^{xy}S_{isw}^{xy} + \sum_{r\neq i}^{g}\sum_{s\neq j}^{m} S_{ijw}^{xy}S_{rsw}^{xy}}{S_w^x} = O_p(1). \tag{2.4.11}$$

Therefore, we obtain

$$S_w^y - \frac{S_w^{xy2}}{S_w^x} = \sum_{i=1}^{g}\sum_{j=1}^{m}(e_{ij}^2 - \dot{\sigma}_e{}^2) + mg\dot{\sigma}_e{}^2 + o_p(g^{\frac{1}{2}}m^{\frac{1}{2}}). \tag{2.4.12}$$

Next, we obtain an approximation for $S_b^y - \frac{(S_b^{xy})^2}{S_b^x}$. For S_b^y and S_b^{xy}, we write

$$
\begin{aligned}
S_b^y &= m\sum_{i=1}^{g}(\bar{y}_{i.} - \bar{y}_{..})^2 \\
&= m\sum_{i=1}^{g}\Big\{\dot{\beta}_1(\bar{x}_{i.} - \bar{x}..) + (\alpha_i - \bar{\alpha}) + (\bar{e}_{i.} - \bar{e}_{..})\Big\}^2 \\
&= m\sum_{i=1}^{g}\dot{\beta}_1^2(\bar{x}_{i.} - \bar{x}..)^2 + m\sum_{i=1}^{g}(\alpha_i - \bar{\alpha})^2 + m\sum_{i=1}^{g}(\bar{e}_{i.} - \bar{e}_{..})^2 + 2m\sum_{i=1}^{g}\dot{\beta}_1(\bar{x}_{i.} - \bar{x}..)(\alpha_i - \bar{\alpha}) \\
&\quad + 2m\sum_{i=1}^{g}\dot{\beta}_1(\bar{x}_{i.} - \bar{x}..)(\bar{e}_{i.} - \bar{e}_{..}) + 2m\sum_{i=1}^{g}(\alpha_i - \bar{\alpha})(\bar{e}_{i.} - \bar{e}_{..}) \\
&= S_b^x\dot{\beta}_1^2 + m\sum_{i=1}^{g}(\alpha_i^2 - 2\alpha_i\bar{\alpha} + \bar{\alpha}^2) + m\sum_{i=1}^{g}(\bar{e}_{i.}^2 - 2\bar{e}_{i.}\bar{e}_{..} + \bar{e}_{..}^2) \\
&\quad + 2m\sum_{i=1}^{g}\dot{\beta}_1\alpha_i + 2m\sum_{i=1}^{g}\dot{\beta}_1(\bar{x}_{i.} - \bar{x}..)\bar{e}_{i.} + 2m\sum_{i=1}^{g}(\alpha_i - \bar{\alpha})\bar{e}_{i.} \\
&= S_b^x\dot{\beta}_1^2 + m\sum_{i=1}^{g}(\alpha_i^2 - \bar{\alpha}^2) + m\sum_{i=1}^{g}(\bar{e}_{i.}^2 - \bar{e}_{..}^2) + 2m\sum_{i=1}^{g}(\bar{x}_{i.} - \bar{x}..)\dot{\beta}_1\alpha_i \\
&\quad + 2m\sum_{i=1}^{g}\dot{\beta}_1(\bar{x}_{i.} - \bar{x}..)\bar{e}_{i.} + 2m\sum_{i=1}^{g}(\alpha_i - \bar{\alpha})\bar{e}_{i.}
\end{aligned}
$$

and

$$
\begin{aligned}
S_b^{xy} &= m\sum_{i=1}^{g}(\bar{y}_{i.} - \bar{y}_{..})(\bar{x}_{i.} - \bar{x}..) \\
&= m\sum_{i=1}^{g}\Big\{\dot{\beta}_1(\bar{x}_{i.} - \bar{x}..)^2 + (\alpha_i - \bar{\alpha})(\bar{x}_{i.} - \bar{x}..) + (\bar{e}_{i.} - \bar{e}_{..})(\bar{x}_{i.} - \bar{x}..)\Big\} \\
&= m\sum_{i=1}^{g}\Big\{\dot{\beta}_1(\bar{x}_{i.} - \bar{x}..)^2 + \alpha_i(\bar{x}_{i.} - \bar{x}..) + \bar{e}_{i.}(\bar{x}_{i.} - \bar{x}..)\Big\} \\
&= S_b^x\dot{\beta}_1 + m\sum_{i=1}^{g}(\bar{x}_{i.} - \bar{x}..)(\alpha_i + \bar{e}_{i.}).
\end{aligned}
$$

On combining the above equations, we obtain

$$
\begin{aligned}
&S_b^y - \frac{(S_b^{xy})^2}{S_b^x} \\
&= \sum_{i=1}^{g}m(\alpha_i^2 - \bar{\alpha}^2) + \sum_{i=1}^{g}m(\bar{e}_{i.}^2 - \bar{e}_{..}^2) + 2\sum_{i=1}^{g}m(\alpha_i - \bar{\alpha})\bar{e}_{i.} \\
&\quad - \frac{\{\sum_{i=1}^{g}m(\bar{x}_{i.} - \bar{x}..)\alpha_i\}^2}{S_b^x} - \frac{2\sum_{i=1}^{g}\sum_{k=1}^{g}m^2(\bar{x}_{i.} - \bar{x}..)(\bar{x}_{k.} - \bar{x}..)\alpha_i\bar{e}_{k.}}{S_b^x} \\
&\quad - \frac{\{\sum_{i=1}^{g}m(\bar{x}_{i.} - \bar{x}..)\bar{e}_{i.}\}^2}{S_b^x}
\end{aligned}
\tag{2.4.13}
$$

and

$$\mathrm{E}\left\{S_b^y - \frac{(S_b^{xy})^2}{S_b^x}\right\} = mg\dot{\sigma_\alpha}^2 - 2m\dot{\sigma_\alpha}^2 + g\dot{\sigma_e}^2 - 2\dot{\sigma_e}^2. \tag{2.4.14}$$

Let $v_i = (\bar{x}_{i.} - \bar{x}..)\alpha_i$ and $w_i = (\bar{x}_{i.} - \bar{x}..)\bar{e}_{i.}$, such that we can rewrite (2.4.13) as

$$\begin{aligned}
&S_b^y - \frac{(S_b^{xy})^2}{S_b^x} \\
&= \sum_{i=1}^g m(\alpha_i^2 - \bar{\alpha}^2) + \sum_{i=1}^g m(\bar{e}_{i.}^2 - \bar{e}_{..}^2) + 2\sum_{i=1}^g m(\alpha_i - \bar{\alpha})\bar{e}_{i.} \\
&\quad - \frac{1}{S_b^x}\left\{(\sum_{i=1}^g mv_i)^2 + 2\sum_{i=1}^g\sum_{k=1}^g m^2 v_i w_k + (\sum_{i=1}^g mw_i)^2\right\} \\
&= \sum_{i=1}^g \left\{m(\alpha_i^2 - \bar{\alpha}^2) + m(\bar{e}_{i.}^2 - \bar{e}_{..}^2) + m(\alpha_i - \bar{\alpha})\bar{e}_{i.}\right\} \\
&\quad - \sum_{i=1}^g \frac{m^2}{S_b^x}(v_i^2 + \sum_{k\neq i}^g v_i v_k + 2v_i w_i + 2\sum_{k\neq i}^g v_i w_k + w_i^2 + \sum_{k\neq i}^g w_i w_k) \\
&= \sum_{i=1}^g \left\{mS_{ib}^y - \frac{m^2(S_{ib}^{xy})^2 + m^2\sum_{k\neq i}^g S_{ib}^{xy}S_{kb}^{xy}}{S_b^x}\right\}.
\end{aligned} \tag{2.4.15}$$

We have

$$\begin{aligned}
&\mathrm{E}(v_i) = 0, \qquad \mathrm{E}(v_i^2) = (\bar{x}_{i.} - \bar{x}..)^2\dot{\sigma_\alpha}^2, \qquad \mathrm{E}(v_i^4) = (\bar{x}_{i.} - \bar{x}..)^4\,\mathrm{E}(\alpha_i^4), \\
&\mathrm{E}(w_i) = 0, \qquad \mathrm{E}(w_i^2) = (\bar{x}_{i.} - \bar{x}..)^2\frac{\dot{\sigma_e}^2}{m} \\
&\text{and} \quad \mathrm{E}(w_i^4) = (\bar{x}_{i.} - \bar{x}..)^4\{m^{-3}\,\mathrm{E}(e_{ij}^4) + 3m^{-2}(1 - m^{-1})\dot{\sigma_e}^4\}.
\end{aligned}$$

By Condition A3, we have that $\lim_{g\to\infty}\frac{1}{g}\sum_{i=1}^g \bar{x}_{i.}^k \to c_k$ for $k = 1, 2, 3, 4$, and $\lim_{g,m\to\infty}(\bar{x}_{i.} - \bar{x}..)^2 < \infty$ and $\lim_{g,m\to\infty}(\bar{x}_{i.} - \bar{x}..)^4 < \infty$. Thus we have

$$mS_{ib}^y - \frac{m^2(S_{ib}^{xy})^2 + m^2\sum_{k\neq i}^g S_{ib}^{xy}S_{kb}^{xy}}{S_b^x} = O_p(m). \tag{2.4.16}$$

Using Lemma1, we can then calculate the variance of each term in (2.4.13),

$$\mathrm{Var}(m\sum_{i=1}^g \alpha_i^2) = m^2\sum_{i=1}^g \mathrm{Var}(\alpha_i^2) = m^2\sum_{i=1}^g\left\{\mathrm{E}(\alpha_i^4) - \dot{\sigma_\alpha}^4\right\} = O(gm^2),$$

$$\begin{aligned}
\operatorname{Var}(m\sum_{i=1}^{g}\bar{\alpha}^2) &= m^2 g^{-2}\operatorname{Var}\left\{(\sum_{i=1}^{g}\alpha_i)^2\right\} \\
&= m^2 g^{-2}\left\{\mathrm{E}(\sum_{i=1}^{g}\sum_{k=1}^{g}\alpha_i\alpha_k)^2 - \mathrm{E}^2(\sum_{i=1}^{g}\sum_{k=1}^{g}\alpha_i\alpha_k)\right\} \\
&= m^2 g^{-2}\sum_{i=1}^{g}\sum_{k=1}^{g}\sum_{r=1}^{g}\sum_{s=1}^{g}\mathrm{E}(\alpha_i\alpha_k\alpha_r\alpha_s) - m^2\dot{\sigma_\alpha}^4 \\
&= m^2 g^{-1}\,\mathrm{E}\,\alpha_i^4 + m^2 g^{-2} 3g(g-1)\dot{\sigma_\alpha}^4 - m^2\dot{\sigma_\alpha}^4 \\
&= O(m^2),
\end{aligned}$$

$$\operatorname{Var}(m\sum_{i=1}^{g}\bar{e}_{i.}^2) = \sum_{i=1}^{g}\operatorname{Var}(m\bar{e}_{i.}^2) = m^{-2}\sum_{i=1}^{g}\sum_{j=1}^{m}\mathrm{E}(e_{ij}^4) + g(2-3m^{-1})\dot{\sigma_e}^4 = O(g),$$

$$\begin{aligned}
\operatorname{Var}(gm\bar{e}_{..}^2) &= g^{-2}m^2\operatorname{Var}\left\{(\sum_{i=1}^{g}\bar{e}_{i.})^2\right\} \\
&= g^{-2}m^2\left\{\sum_{i=1}^{g}\sum_{k=1}^{g}\sum_{r=1}^{g}\sum_{s=1}^{g}\mathrm{E}\,\bar{e}_{i.}\bar{e}_{k.}\bar{e}_{r.}\bar{e}_{s.} - (\sum_{i=1}^{g}\sum_{k=1}^{g}\mathrm{E}\,\bar{e}_{i.}\bar{e}_{k.})^2\right\} \\
&= g^{-2}m^2\left\{\sum_{i=1}^{g}\sum_{k=1}^{g}\sum_{r=1}^{g}\sum_{s=1}^{g}\mathrm{E}\,\bar{e}_{i.}\bar{e}_{k.}\bar{e}_{r.}\bar{e}_{s.} - (\sum_{i=1}^{g}\mathrm{E}\,\bar{e}_{i.}^2)^2\right\} \\
&= g^{-2}m^2\left\{\sum_{i=1}^{g}\mathrm{E}(\bar{e}_{i.}^4) + 3\sum_{i=1}^{g}\sum_{k\neq i}^{g}\mathrm{E}(\bar{e}_{i.}^2)\,\mathrm{E}(\bar{e}_{k.}^2) - (\sum_{i=1}^{g}\mathrm{E}\,\bar{e}_{i.}^2)^2\right\} \\
&= g^{-2}m^2\left\{m^{-4}\sum_{i=1}^{g}\sum_{j=1}^{m}\mathrm{E}(e_{ij}^4) + 3m^{-2}(1-m^{-1})\dot{\sigma_e}^4 + 3g(g-1)m^{-2}\dot{\sigma_e}^2 - g^2m^{-2}\dot{\sigma_e}^2\right\} \\
&= O(1),
\end{aligned}$$

$$\operatorname{Var}(2m\sum_{i=1}^{g}\alpha_i\bar{e}_{i.}) = 4m^2\sum_{i=1}^{g}\left\{\mathrm{E}\,\alpha_i^2\,\mathrm{E}\,\bar{e}_{i.}^2 - (\mathrm{E}\,\alpha_i\,\mathrm{E}\,\bar{e}_{i.})^2\right\} = 4m^2\sum_{i=1}^{g}\mathrm{E}\,\alpha_i^2\,\mathrm{E}\,\bar{e}_{i.}^2 = 4mg\dot{\sigma_\alpha}^2\dot{\sigma_e}^2 = O(gm),$$

$$\operatorname{Var}(m\sum_{i=1}^{g}\bar{\alpha}\bar{e}_{i.}) = m^2\sum_{i=1}^{g}\operatorname{Var}(\bar{\alpha}\bar{e}_{i.}) = m^2\sum_{i=1}^{g}\mathrm{E}(\bar{\alpha}^2)\,\mathrm{E}(\bar{e}_{i.}^2) = m^2gg^{-1}\dot{\sigma_\alpha}^2m^{-1}\dot{\sigma_e}^2 = m\dot{\sigma_\alpha}^2\dot{\sigma_e}^2 = O(m),$$

$$
\begin{aligned}
&\operatorname{Var}\left[\frac{\{\sum_{i=1}^{g} m(\bar{x}_{i.}-\bar{x}_{..})\alpha_i\}^2}{S_b^x}\right] \\
&= \frac{m^4}{(S_b^x)^2}\operatorname{Var}\left\{(\sum_{i=1}^{g} v_i)^2\right\} \\
&= \frac{m^4}{(S_b^x)^2}\left[\sum_{i=1}^{g}\sum_{k=1}^{g}\sum_{r=1}^{g}\sum_{s=1}^{g}\mathrm{E}(v_i v_k v_r v_s) - \left\{\sum_{i=1}^{g}\sum_{k=1}^{g}\mathrm{E}(v_i v_k)\right\}^2\right] \\
&= \frac{m^4}{(S_b^x)^2}\left\{\sum_{i=1}^{g}\mathrm{E}(v_i^4) - 3\sum_{i=1}^{g}\sum_{k\neq i}\mathrm{E}(v_i^2)\,\mathrm{E}(v_k^2) - \sum_{i=1}^{g}\sum_{k=1}^{g}\mathrm{E}(v_i^2)\,\mathrm{E}(v_k^2)\right\} \\
&= O(g^{-1}m^2) + O(m^2) + O(m^2) \\
&= O(m^2),
\end{aligned}
$$

$$
\begin{aligned}
&\operatorname{Var}\left\{\frac{\sum_{i=1}^{g}\sum_{k=1}^{g} m^2(\bar{x}_{i.}-\bar{x}_{..})(\bar{x}_{k.}-\bar{x}_{..})\alpha_i\bar{e}_{k.}}{S_b^x}\right\} \\
&= \frac{m^4}{(S_b^x)^2}\operatorname{Var}\left(\sum_{i=1}^{g}\sum_{k=1}^{g} v_i w_k\right) \\
&= \frac{m^4}{(S_b^x)^2}\left[\mathrm{E}\left\{(\sum_{i=1}^{g}\sum_{k=1}^{g} v_i w_k)^2\right\} - \left\{\mathrm{E}(\sum_{i=1}^{g}\sum_{k=1}^{g} v_i w_k)\right\}^2\right] \\
&= \frac{m^4}{(S_b^x)^2}\left\{\sum_{i=1}^{g}\sum_{k=1}^{g}\sum_{r=1}^{g}\sum_{s=1}^{g}\mathrm{E}(v_i v_r w_k w_s) - 0\right\} \\
&= \frac{m^4}{(S_b^x)^2}\sum_{i=1}^{g}\sum_{k=1}^{g}\sum_{r=1}^{g}\sum_{s=1}^{g}\mathrm{E}(v_i v_r)\,\mathrm{E}(w_k w_s) \\
&= \frac{m^4}{(S_b^x)^2}\sum_{i=1}^{g}\sum_{k=1}^{g}\mathrm{E}(v_i^2)\,\mathrm{E}(w_k^2) \\
&= O(m)
\end{aligned}
$$

and

$$
\begin{aligned}
&\operatorname{Var}\left[\frac{\{\sum_{i=1}^{g} m(\bar{x}_{i.}-\bar{x}_{..})\bar{e}_{i.}\}^2}{S_b^x}\right] \\
&= \frac{m^4}{(S_b^x)^2}\operatorname{Var}\left\{(\sum_{i=1}^{g} w_i)^2\right\} \\
&= \frac{m^4}{(S_b^x)^2}\left[\sum_{i=1}^{g}\sum_{k=1}^{g}\sum_{r=1}^{g}\sum_{s=1}^{g}\mathrm{E}(w_i w_k w_r w_s) - \left\{\sum_{i=1}^{g}\sum_{k=1}^{g}\mathrm{E}(w_i w_k)\right\}^2\right] \\
&= \frac{m^4}{(S_b^x)^2}\left\{\sum_{i=1}^{g}\mathrm{E}(w_i^4) - 3\sum_{i=1}^{g}\sum_{k\neq i}\mathrm{E}(w_i^2)\,\mathrm{E}(w_k^2) - \sum_{i=1}^{g}\sum_{k=1}^{g}\mathrm{E}(w_i^2)\,\mathrm{E}(w_k^2)\right\} \\
&= O(g^{-2}m^2)\left\{O(gm^{-2}) + O(g^2m^{-1}) + O(g^2m^{-1})\right\} \\
&= O(m).
\end{aligned}
$$

Combining the above equations, we obtain

$$S_b^y - \frac{S_b^{xy2}}{S_b^x} = \sum_{i=1}^{g} m(\alpha_i^2 - \dot{\sigma_\alpha}^2) + mg\dot{\sigma_\alpha}^2 + o_p(g^{\frac{1}{2}}m). \tag{2.4.17}$$

Let $\mathbf{K}_R = \mathrm{diag}[g, gm]$. Using (2.4.12) and (2.4.17), we can finally obtain an approximation the REML score equations as

$$\mathbf{K}_R^{\frac{1}{2}}\psi_R(\boldsymbol{\theta}_0) = \mathbf{K}_R^{\frac{1}{2}}\xi_R(\boldsymbol{\theta}_0) + o_p(1), \tag{2.4.18}$$

where $\xi_R(\boldsymbol{\theta}_0)$ has components,

$$\begin{aligned} \xi_{R\sigma_\alpha^2}(\boldsymbol{\theta}_0) &= \sum_{i=1}^{g} \frac{m^2}{2(\dot{\sigma_e}^2 + m\dot{\sigma_\alpha}^2)^2}(\alpha_i^2 - \dot{\sigma_\alpha}^2), \\ \xi_{R\sigma_e^2}(\boldsymbol{\theta}_0) &= \frac{1}{2\dot{\sigma_e}^4}\sum_{i=1}^{g}\sum_{j=1}^{m}(e_{ij}^2 - \dot{\sigma_e}^2). \end{aligned} \tag{2.4.19}$$

We see that our approximation of REML estimating equation at $\boldsymbol{\theta}_0$ is much simpler, while the estimating equation for REML estimator is a complicated function of the cluster sizes. As $g, m \to \infty$, the form of $\psi_R(\boldsymbol{\theta}_0)$ simplifies and we can develop a simple approximation to $\psi_R(\boldsymbol{\theta}_0)$.

2.4.2 Computing G_n, the variance of the approximate REML estimating equation at $\boldsymbol{\theta}_0$

We calculate the variance of $\mathbf{K}_R^{-\frac{1}{2}}\xi_R(\boldsymbol{\theta}_0)$ as

$$\mathbf{G}_{Rn} = \mathrm{Var}\begin{bmatrix} g^{-\frac{1}{2}}\xi_R^{(b)}(\boldsymbol{\theta}_0) \\ g^{-\frac{1}{2}}m^{-\frac{1}{2}}\xi_R^{(w)}(\boldsymbol{\theta}_0) \end{bmatrix} = \begin{bmatrix} \frac{m^4(\mathrm{E}\,\alpha_i^4 - \dot{\sigma_\alpha}^4)}{4(\dot{\sigma_e}^2 + m\dot{\sigma_\alpha}^2)^4} & 0 \\ 0 & \frac{\mathrm{E}\,e_{ij}^4 - \dot{\sigma_e}^4}{4\dot{\sigma_e}^8} \end{bmatrix}. \tag{2.4.20}$$

Since $\lim_{m\to\infty} \frac{m}{(\dot{\sigma_e}^2 + m\dot{\sigma_\alpha}^2)} = \frac{1}{\dot{\sigma_\alpha}^2}$, then we have

$$\mathbf{G}_R = \lim_{m\to\infty} \mathbf{G}_{Rn} = \begin{bmatrix} \frac{\mathrm{E}\,\alpha_i^4 - \dot{\sigma_\alpha}^4}{4\dot{\sigma_\alpha}^8} & 0 \\ 0 & \frac{\mathrm{E}\,e_{ij}^4 - \dot{\sigma_e}^4}{4\dot{\sigma_e}^8} \end{bmatrix}. \tag{2.4.21}$$

2.4.3 Computing the derivative matrix from the REML estimating equations

From (2.4.7), we can show that

$$\begin{aligned} l_{R\sigma_\alpha^2\sigma_\alpha^2}(\boldsymbol{\theta}) &= \frac{gm^2(1-2g^{-1})}{2(\sigma_e^2 + m\sigma_\alpha^2)^2} - \frac{m^2}{(\sigma_e^2 + m\sigma_\alpha^2)^3}\left\{S_b^y - \frac{(S_b^{xy})^2}{S_b^x}\right\} \\ l_{R\sigma_\alpha^2\sigma_e^2}(\boldsymbol{\theta}) = l_{R\sigma_e^2\sigma_\alpha^2}(\boldsymbol{\theta}) &= \frac{gm(1-2g^{-1})}{2(\sigma_e^2 + m\sigma_\alpha^2)^2} - \frac{m}{(\sigma_e^2 + m\sigma_\alpha^2)^3}\left\{S_b^y - \frac{(S_b^{xy})^2}{S_b^x}\right\} \\ l_{R\sigma_e^2\sigma_e^2}(\boldsymbol{\theta}) &= \frac{gm(1-2g^{-1})}{2(\sigma_e^2 + m\sigma_\alpha^2)^2} + \frac{g(m-1-g^{-1})}{2\sigma_e^4} - \frac{1}{\sigma_e^6}\left\{S_w^y - \frac{(S_w^{xy})^2}{S_w^x}\right\} - \frac{1}{(\sigma_e^2 + m\sigma_\alpha^2)^3}\left\{S_b^y - \frac{(S_b^{xy})^2}{S_b^x}\right\} \end{aligned} \tag{2.4.22}$$

From (2.4.10) and (2.4.15), let $l_{R\sigma_\alpha^2\sigma_\alpha^2}(\boldsymbol{\theta}) = \sum_{i=1}^g l_{iR\sigma_\alpha^2\sigma_\alpha^2}(\boldsymbol{\theta})$, $l_{R\sigma_\alpha^2\sigma_e^2}(\boldsymbol{\theta}) = \sum_{i=1}^g l_{iR\sigma_\alpha^2\sigma_e^2}(\boldsymbol{\theta})$, $l_{R\sigma_e^2\sigma_\alpha^2}(\boldsymbol{\theta}) = \sum_{i=1}^g \sum_{j=1}^m l_{ijR\sigma_e^2\sigma_\alpha^2}(\boldsymbol{\theta})$ and $l_{R\sigma_e^2\sigma_e^2}(\boldsymbol{\theta}) = \sum_{i=1}^g \sum_{j=1}^m l_{ijR\sigma_e^2\sigma_e^2}(\boldsymbol{\theta})$, where

$$\begin{aligned}
l_{iR\sigma_\alpha^2\sigma_\alpha^2}(\boldsymbol{\theta}) &= \frac{m^2(1-2g^{-1})}{2(\sigma_e^2+m\sigma_\alpha^2)^2} - \frac{m^2}{(\sigma_e^2+m\sigma_\alpha^2)^3}\left\{mS_{ib}^y - \frac{m^2(S_{ib}^{xy})^2 + m^2\sum_{k\neq i}^g S_{ib}^{xy}S_{kb}^{xy}}{S_b^x}\right\}\\
l_{iR\sigma_\alpha^2\sigma_e^2}(\boldsymbol{\theta}) &= \frac{m(1-2g^{-1})}{2(\sigma_e^2+m\sigma_\alpha^2)^2} - \frac{m}{(\sigma_e^2+m\sigma_\alpha^2)^3}\left\{mS_{ib}^y - \frac{m^2(S_{ib}^{xy})^2 + m^2\sum_{k\neq i}^g S_{ib}^{xy}S_{kb}^{xy}}{S_b^x}\right\}\\
l_{ijR\sigma_\alpha^2\sigma_e^2}(\boldsymbol{\theta}) &= \frac{(1-2g^{-1})}{2(\sigma_e^2+m\sigma_\alpha^2)^2} - \frac{1}{(\sigma_e^2+m\sigma_\alpha^2)^3}\left\{mS_{ib}^y - \frac{m^2(S_{ib}^{xy})^2 + m^2\sum_{k\neq i}^g S_{ib}^{xy}S_{kb}^{xy}}{S_b^x}\right\}\\
l_{ijR\sigma_e^2\sigma_e^2}(\boldsymbol{\theta}) &= \frac{(1-2g^{-1})}{2(\sigma_e^2+m\sigma_\alpha^2)^2} + \frac{(m-1-g^{-1})}{2m\sigma_e^4}\\
&\quad - \frac{1}{\sigma_e^6}\left\{S_{ijw}^x - \frac{(S_{ijw}^{xy})^2 + \sum_{r\neq i}^g S_{ijw}^{xy}S_{rjw}^{xy} + \sum_{s\neq j}^m S_{ijw}^{xy}S_{isw}^{xy} + \sum_{r\neq i}^g\sum_{s\neq j}^m S_{ijw}^{xy}S_{rsw}^{xy}}{S_w^x}\right\}\\
&\quad - \frac{1}{(\sigma_e^2+m\sigma_\alpha^2)^3}\left\{S_{ib}^y - \frac{m(S_{ib}^{xy})^2 + m\sum_{k\neq i}^g S_{ib}^{xy}S_{kb}^{xy}}{S_b^x}\right\}.
\end{aligned} \tag{2.4.23}$$

Based on (2.4.11) and (2.4.16) , we can show that

$$l_{iR\sigma_\alpha^2\sigma_\alpha^2}(\boldsymbol{\theta}) = O_p(1), \quad l_{iR\sigma_\alpha^2\sigma_e^2}(\boldsymbol{\theta}) = O_p(m^{-1}) \quad l_{ijR\sigma_\alpha^2\sigma_e^2}(\boldsymbol{\theta}) = O_p(m^{-2}) \quad \text{and} \quad l_{ijR\sigma_e^2\sigma_e^2}(\boldsymbol{\theta}) = O_p(1). \tag{2.4.24}$$

We now proceed to calculate the expected derivative matrix

$$\mathrm{E}\left\{\nabla\psi_R(\boldsymbol{\theta}\right\} = \begin{bmatrix} \sum_{i=1}^g \mathrm{E}\left\{l_{iR\sigma_\alpha^2\sigma_\alpha^2}(\boldsymbol{\theta}\right\} & \sum_{i=1}^g \mathrm{E}\left\{l_{iR\sigma_\alpha^2\sigma_e^2}(\boldsymbol{\theta}\right\} \\ \sum_{i=1}^g\sum_{j=1}^m \mathrm{E}\left\{l_{ijR\sigma_e^2\sigma_\alpha^2}(\boldsymbol{\theta}\right\} & \sum_{i=1}^g\sum_{j=1}^m \mathrm{E}\left\{l_{ijR\sigma_e^2\sigma_e^2}(\boldsymbol{\theta}\right\}\end{bmatrix},$$

where

$$\begin{aligned}
\mathrm{E}\left\{l_{iR\sigma_\alpha^2\sigma_\alpha^2}(\boldsymbol{\theta}\right\} &= \frac{m^2(1-2g^{-1})}{2(\sigma_e^2+m\sigma_\alpha^2)^2} - \frac{m^2(m\dot{\sigma_\alpha}^2 - 2g^{-1}m\dot{\sigma_\alpha}^2 + \dot{\sigma_e}^2 - 2g^{-1}\dot{\sigma_e}^2)}{(\sigma_e^2+m\sigma_\alpha^2)^3} = O(1)\\
\mathrm{E}\left\{l_{iR\sigma_\alpha^2\sigma_e^2}(\boldsymbol{\theta}\right\} &= \frac{m(1-2g^{-1})}{2(\sigma_e^2+m\sigma_\alpha^2)^2} - \frac{m(m\dot{\sigma_\alpha}^2 - 2g^{-1}m\dot{\sigma_\alpha}^2 + \dot{\sigma_e}^2 - 2g^{-1}\dot{\sigma_e}^2)}{(\sigma_e^2+m\sigma_\alpha^2)^3} = O(m^{-1})\\
\mathrm{E}\left\{l_{ijR\sigma_e^2\sigma_\alpha^2}(\boldsymbol{\theta}\right\} &= \frac{1-2g^{-1}}{2(\sigma_e^2+m\sigma_\alpha^2)^2} - \frac{(m\dot{\sigma_\alpha}^2 - 2g^{-1}m\dot{\sigma_\alpha}^2 + \dot{\sigma_e}^2 - 2g^{-1}\dot{\sigma_e}^2)}{(\sigma_e^2+m\sigma_\alpha^2)^3} = O(m^{-2})\\
\mathrm{E}\left\{l_{ijR\sigma_e^2\sigma_e^2}(\boldsymbol{\theta})\right\} &= \frac{1-2g^{-1}}{2m(\sigma_e^2+m\sigma_\alpha^2)^2} + \frac{(m-1-g^{-1})}{2m\sigma_e^4} - \frac{\dot{\sigma_e}^2 - (g^{-1}m^{-1}+m^{-1})\dot{\sigma_e}^2}{\sigma_e^6}\\
&\quad - \frac{\dot{\sigma_\alpha}^2 - 2g^{-1}\dot{\sigma_\alpha}^2 + m^{-1}\dot{\sigma_e}^2 - 2g^{-1}m^{-1}\dot{\sigma_e}^2}{(\sigma_e^2+m\sigma_\alpha^2)^3} = O(1).
\end{aligned} \tag{2.4.25}$$

When $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, we have,

$$\begin{aligned}
\mathrm{E}\left\{l_{iR\sigma_\alpha^2\sigma_\alpha^2}(\boldsymbol{\theta}_0)\right\} &= -\frac{m^2(1-2g^{-1})}{2(\dot{\sigma_e}^2 + m\dot{\sigma_\alpha}^2)^2} = O(1) \\
\mathrm{E}\left\{l_{iR\sigma_\alpha^2\sigma_e^2}(\boldsymbol{\theta}_0)\right\} &= -\frac{m(1-2g^{-1})}{2(\dot{\sigma_e}^2 + m\dot{\sigma_\alpha}^2)^2} = O(m^{-1}) \\
\mathrm{E}\left\{l_{ijR\sigma_e^2\sigma_\alpha^2}(\boldsymbol{\theta}_0)\right\} &= -\frac{(1-2g^{-1})}{2(\dot{\sigma_e}^2 + m\dot{\sigma_\alpha}^2)^2} = O(m^{-2}) \\
\mathrm{E}\left\{l_{ijR\sigma_e^2\sigma_e^2}(\boldsymbol{\theta}_0)\right\} &= -\frac{1-2g^{-1}}{2(\dot{\sigma_e}^2 + m\dot{\sigma_\alpha}^2)^2} - \frac{m-1-g^{-1}}{2m\dot{\sigma_e}^4} = O(1).
\end{aligned} \tag{2.4.26}$$

Thus, we can write

$$\mathbf{B}_{Rn} = -\mathbf{K}_R^{-\frac{1}{2}}\,\mathrm{E}\,\nabla\psi_R(\boldsymbol{\theta}_0)\mathbf{K}_R^{-\frac{1}{2}} = \begin{bmatrix} \frac{m^2(1-2g^{-1})}{2(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)^2} & \frac{m(1-2g^{-1})}{2(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)^2} \\ \frac{(1-2g^{-1})}{2(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)^2} & \frac{1-2g^{-1}}{2m(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)^2} + \frac{m-1-g^{-1}}{2m\dot{\sigma_e}^4} \end{bmatrix}$$

Since $\lim_{m\to\infty} \frac{m}{(\dot{\sigma_e}^2+m\dot{\sigma_\alpha}^2)} = \frac{1}{\dot{\sigma_\alpha}^2}$, we have

$$\mathbf{B}_R = \lim_{g,m\to\infty} \mathbf{B}_{Rn} = \begin{bmatrix} \frac{1}{2\dot{\sigma_\alpha}^4} & 0 \\ 0 & \frac{1}{2\dot{\sigma_e}^4} \end{bmatrix}. \tag{2.4.27}$$

From (2.4.21) and (2.4.21), we can see that under normality, $\mathbf{G}_R = \mathbf{B}_R$.

2.5 Main results on REML estimators

The first step in establishing a central limit theorem for the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ is to find an appropriate normalizing matrix $\mathbf{K}$ that ensures that $\mathbf{K}_R^{-\frac{1}{2}}\{\psi_R(\boldsymbol{\theta}_0) - \xi(\boldsymbol{\theta}_0)\} = o_p(1)$ and that $\mathbf{K}_R^{-\frac{1}{2}}\xi_R(\boldsymbol{\theta}_0) \xrightarrow{D} N(\mathbf{0}, \mathbf{G}_R)$, where $\mathbf{G}_R$ is a positive definite covariance matrix. That is, $\mathbf{K}_R^{-\frac{1}{2}}\psi_R(\boldsymbol{\theta}_0)$ can be approximated by $\mathbf{K}_R^{-\frac{1}{2}}\xi_R(\boldsymbol{\theta}_0)$ which converges in distribution to a $N(\mathbf{0}, \mathbf{G}_R)$ distribution. Under condition A, we show that the appropriate normalizing matrix is the diagonal matrix $\mathbf{K}_R = \mathrm{diag}[g, mg]$. This justifies focusing our attention on the neighborhood of $\boldsymbol{\theta}_0$ defined by $\mathcal{B} = \{\boldsymbol{\theta} : |\mathbf{K}_R^{\frac{1}{2}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)| \leq M\}$, where M is a finite constant. The second step is to show that, uniformly on $\mathcal{B}$, the derivative matrix $\nabla\psi_R(\boldsymbol{\theta})$ of $\psi_R(\boldsymbol{\theta})$, pre- and post-multiplied by $\mathbf{K}_R^{-\frac{1}{2}}$, converges in probability to a positive definite matrix $-\mathbf{B}_R$. We can then establish that, uniformly on $\mathcal{B}$,

$$\mathbf{K}^{-\frac{1}{2}}\psi(\boldsymbol{\omega}) = \mathbf{K}^{-\frac{1}{2}}\xi(\boldsymbol{\omega}) - \mathbf{B}\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega} - \boldsymbol{\omega}_0) + o_p(1).$$

This uniform asymptotic linearity result allows us to deduce the existence of a consistent root to the estimating equation that satisfies the central limit theorem with asymptotic variance.

Theorem 2. *Suppose Condition A holds. Then there is a solution $\hat{\boldsymbol{\theta}}$ to the estimating equations $\psi_R(\boldsymbol{\theta}) = 0$, satisfying $|\mathbf{K}_R^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)| = O_p(1)$ as $g \to \infty$ and $m \to \infty$, where $\mathbf{K}_R = \mathrm{diag}[g, mg]$. Moreover,*

$$\mathbf{K}_R^{\frac{1}{2}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \mathbf{B}_R^{-1}\mathbf{K}_R^{-\frac{1}{2}}\xi_R(\boldsymbol{\theta}_0) + o_p(1),$$

whence

$$\mathbf{K}_R^{\frac{1}{2}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{D}, N(\mathbf{0}, \mathbf{B}_R^{-1}\mathbf{G}_R\mathbf{B}_R^{-1}),$$

where $\mathbf{B}_R = \lim_{g,m\to\infty}\mathbf{B}_{Rn} = -\mathbf{K}_R^{-\frac{1}{2}}\,\mathrm{E}\left\{\nabla\psi_R(\boldsymbol{\theta}_0)\right\}\mathbf{K}_R^{-\frac{1}{2}}$ *and* $\mathbf{G}_R = \lim_{g,m\to\infty}\mathbf{G}_{Rn} = \mathrm{Var}\left\{\mathbf{K}_R^{-\frac{1}{2}}\xi_R(\boldsymbol{\theta}_0)\right\}$.

To simplify the proofs, we establish a preliminary lemma and then break the main proof into a sequence of lemmas. In order to getting results, we need to use following lemmas.

Lemma 9. *Suppose Condition A holds. Then if $g \to \infty$ and $m \to \infty$, we have $\mathbf{G}_{Rn}^{-\frac{1}{2}}\mathbf{K}_R^{-\frac{1}{2}}\xi_R(\boldsymbol{\theta}_0) \xrightarrow{D}$ $N(\mathbf{0}, \mathbf{I})$.*

Proof. The first step is to prove $\left[\mathrm{Var}\left\{g^{-\frac{1}{2}}\xi_R^{(b)}(\boldsymbol{\theta}_0)\right\}\right]^{-\frac{1}{2}} g^{-\frac{1}{2}}\xi_R^{(b)}(\boldsymbol{\theta}_0) \xrightarrow{D} N(0,1)$. From (2.4.19) and (2.4.20), we write

$$\left[\mathrm{Var}\left\{g^{-\frac{1}{2}}\xi_R^{(b)}(\boldsymbol{\theta}_0)\right\}\right]^{-\frac{1}{2}} g^{-\frac{1}{2}}\xi_R^{(b)}(\boldsymbol{\theta}_0) = \sum_{i=1}^{g} g^{-\frac{1}{2}}\left\{\mathrm{E}(\alpha_i^4) - \dot{\sigma_\alpha}^4\right\}^{-\frac{1}{2}}(\alpha_i^2 - \dot{\sigma_\alpha}^2).$$

Since $\mathrm{E}(\alpha_i^{4+\lambda}) < \infty$, there exist a $\delta > 0$ such that

$$\lim_{g\to\infty}\sum_{i=1}^{g}\mathrm{E}\,|\left\{\mathrm{E}(\alpha_i^4) - \dot{\sigma_\alpha}^4\right\}^{-\frac{1}{2}} g^{-\frac{1}{2}}(\alpha_i^2 - \dot{\sigma_\alpha}^2)|^{2+\delta} = \lim_{g\to\infty} g^{-1-\frac{\delta}{2}}\left\{\mathrm{E}(\alpha_i^4) - \dot{\sigma_\alpha}^4\right\}^{-1-\frac{\delta}{2}}\sum_{i=1}^{g}\mathrm{E}(\alpha_i^2 - \dot{\sigma_\alpha}^2)^{2+\delta} = o_p(1).$$

We find that Lyapunov's condition holds.

The second step is to prove $\left[\mathrm{Var}\left\{g^{-\frac{1}{2}}m^{-\frac{1}{2}}\xi_R^{(w)}(\boldsymbol{\theta}_0)\right\}\right]^{-\frac{1}{2}} g^{-\frac{1}{2}}m^{-\frac{1}{2}}\xi_R^{(w)}(\boldsymbol{\theta}_0) \xrightarrow{D} N(0,1)$. From (2.4.19) and (2.4.20), we write

$$\left[\mathrm{Var}\left\{g^{-\frac{1}{2}}m^{-\frac{1}{2}}\xi_R^{(w)}(\boldsymbol{\theta}_0)\right\}\right]^{-\frac{1}{2}} g^{-\frac{1}{2}}m^{-\frac{1}{2}}\xi_R^{(w)}(\boldsymbol{\theta}_0) = \sum_{i=1}^{g}\sum_{j=1}^{m}(\mathrm{E}\,e_{ij}^4 - \dot{\sigma_e}^4)^{-\frac{1}{2}} g^{-\frac{1}{2}}m^{-\frac{1}{2}}(e_{ij}^2 - \dot{\sigma_e}^2).$$

Since $\mathrm{E}(e_{ij}^{4+\lambda}) < \infty$, there exist a $\delta > 0$ such that

$$\begin{aligned}
&\lim_{g,m\to\infty} \sum_{i=1}^{g}\sum_{j=1}^{m} \mathrm{E}\,|(\mathrm{E}\,e_{ij}^4 - \dot{\sigma_e}^4)^{-\frac{1}{2}} g^{-\frac{1}{2}} m^{-\frac{1}{2}} (e_{ij}^2 - \dot{\sigma_e}^2)|^{2+\delta} \\
&= \lim_{g,m\to\infty} g^{-1-\frac{\delta}{2}} m^{-1-\frac{\delta}{2}} (\mathrm{E}\,e_{ij}^4 - \dot{\sigma_e}^4)^{-1-\frac{\delta}{2}} \sum_{i=1}^{g}\sum_{j=1}^{m} \mathrm{E}(e_{ij}^2 - \dot{\sigma_e}^2)^{2+\delta} \\
&= o_p(1).
\end{aligned}$$

We find that Lyapunov's condition holds. □

Lemma 10. *Suppose Condition A holds. Then, as $g, m \to \infty$,*

$$\sup_{|\mathbf{K}_R^{\frac{1}{2}}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)|\leq M} |\mathbf{K}_R^{-\frac{1}{2}} \left[\psi_R(\boldsymbol{\theta}) - \psi_R(\boldsymbol{\theta}_0) - \mathrm{E}\left\{\psi_R(\boldsymbol{\theta}) - \psi_R(\boldsymbol{\theta}_0)\right\}\right]| = o_p(1).$$

Proof. Decompose the cube $\mathcal{B} = \{\boldsymbol{\theta} : |\mathbf{K}_R^{\frac{1}{2}}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)| \leq M\}$ into smaller cubes $\mathcal{C} = \{\mathcal{C}(t_k)\}$, where M is a constant, and $\{\mathcal{C}(t_k)\}$ is a cube containing $\boldsymbol{\theta} \in \mathcal{C}(t_k) = \{\boldsymbol{\theta} : |\mathbf{K}_R^{\frac{1}{2}}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)| \leq Mg^{-\frac{1}{4}}\}$. Let $N = g^{\frac{1}{4}}$. We first show that lemma 10 holds over the $\mathcal{C}$ of fixed $t_k = (t_{k1}, t_{k2})^T$. Using Chebyshev's inequality, for any $\eta > 0$, we have

$$\begin{aligned}
&\sum_{k=1}^{N} \Pr\left(|\mathbf{K}_R^{-\frac{1}{2}} \left[\psi_R(t_k) - \psi_R(\boldsymbol{\theta}_0) - \mathrm{E}\left\{\psi_R(t_k) - \psi_R(\boldsymbol{\theta}_0)\right\}\right]| > \eta\right) \\
&\leq \eta^{-2} \sum_{k=1}^{N} \mathrm{E}\,|\mathbf{K}_R^{-\frac{1}{2}} \left[\psi_R(t_k) - \psi_R(\boldsymbol{\theta}_0) - \mathrm{E}\left\{\psi_R(t_k) - \psi_R(\boldsymbol{\theta}_0)\right\}\right]|^2 \\
&= \eta^{-2} \sum_{k=1}^{N} \mathrm{E}\,|g^{-\frac{1}{2}} \left[l_{R\sigma_\alpha^2}(t_k) - l_{R\sigma_\alpha^2}(\boldsymbol{\theta}_0) - \mathrm{E}\left\{l_{R\sigma_\alpha^2}(t_k) - l_{R\sigma_\alpha^2}(\boldsymbol{\theta}_0)\right\}\right]|^2 \\
&\quad + \eta^{-2} \sum_{k=1}^{N} \mathrm{E}\,|g^{-\frac{1}{2}} m^{-\frac{1}{2}} \left[l_{R\sigma_e^2}(t_k) - l_{R\sigma_e^2}(\boldsymbol{\theta}_0) - \mathrm{E}\left\{l_{R\sigma_e^2}(t_k) - l_{R\sigma_e^2}(\boldsymbol{\theta}_0)\right\}\right]|^2 \\
&= \eta^{-2} g^{-1} \sum_{k=1}^{N} \mathrm{E}\left[|l_{R\sigma_\alpha^2}(t_k) - l_{R\sigma_\alpha^2}(\boldsymbol{\theta}_0) - \mathrm{E}\left\{l_{R\sigma_\alpha^2}(t_k) - l_{R\sigma_\alpha^2}(\boldsymbol{\theta}_0)\right\}|^2\right] \\
&\quad + \eta^{-2} g^{-1} m^{-1} \sum_{k=1}^{N} \mathrm{E}\left[|l_{R\sigma_e^2}(t_k) - l_{R\sigma_e^2}(\boldsymbol{\theta}_0) - \mathrm{E}\left\{l_{R\sigma_e^2}(t_k) - l_{R\sigma_e^2}(\boldsymbol{\theta}_0)\right\}|^2\right].
\end{aligned} \tag{2.5.1}$$

For the first term of(2.5.1), Since $|t_{k1} - \dot{\sigma_\alpha}^2| \leq g^{-\frac{1}{2}}L$ and $|t_{k2} - \dot{\sigma_e}^2| \leq g^{-\frac{1}{2}}L$, then using Lemma

2 and Taylor expansion, we obtain

$$
\begin{aligned}
&\eta^{-2}g^{-1}\sum_{k=1}^{N}\mathrm{E}\left[|l_{R\sigma_\alpha^2}(t_k)-l_{R\sigma_\alpha^2}(\boldsymbol{\theta}_0)-\mathrm{E}\left\{l_{R\sigma_\alpha^2}(t_k)-l_{R\sigma_\alpha^2}(\boldsymbol{\theta}_0)\right\}|^2\right]\\
&\leq \eta^{-2}g^{-1}\sum_{k=1}^{N}\mathrm{E}\left\{|l_{R\sigma_\alpha^2}(t_k)-l_{R\sigma_\alpha^2}(\boldsymbol{\theta}_0)|^2\right\}\\
&= \eta^{-2}g^{-1}\sum_{k=1}^{N}\mathrm{E}\left[|\sum_{i=1}^{g}\left\{l_{iR\sigma_\alpha^2}(t_k)-l_{iR\sigma_\alpha^2}(\boldsymbol{\theta}_0)\right\}|^2\right]\\
&\leq \eta^{-2}g^{-1}\sum_{k=1}^{N}\sum_{i=1}^{g}\mathrm{E}\left\{|l_{iR\sigma_\alpha^2}(t_k)-l_{iR\sigma_\alpha^2}(\boldsymbol{\theta}_0)|^2\right\}\\
&= \eta^{-2}g^{-1}\sum_{k=1}^{N}\sum_{i=1}^{g}\mathrm{E}\left\{|\frac{\partial l_{iR\sigma_\alpha^2}(\tilde{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}}(t_k-\boldsymbol{\theta}_0)|^2\right\}\\
&= \eta^{-2}g^{-1}\sum_{k=1}^{N}\sum_{i=1}^{g}\mathrm{E}\,|l_{iR\sigma_\alpha^2\sigma_\alpha^2}(\tilde{\boldsymbol{\theta}})(t_{k1}-\dot{\sigma_\alpha}^2)|^2+\eta^{-2}g^{-1}\sum_{k=1}^{N}\sum_{i=1}^{g}\mathrm{E}\,|l_{iR\sigma_\alpha^2\sigma_e^2}(\tilde{\boldsymbol{\theta}})(t_{k2}-\dot{\sigma_e}^2)|^2\\
&\leq \eta^{-2}g^{-2}L\sum_{k=1}^{N}\sum_{i=1}^{g}\mathrm{E}\left\{l^2_{iR\sigma_\alpha^2\sigma_\alpha^2}(\tilde{\boldsymbol{\theta}})\right\}+\eta^{-2}g^{-2}L\sum_{k=1}^{N}\sum_{i=1}^{g}\mathrm{E}\left\{l^2_{iR\sigma_\alpha^2\sigma_e^2}(\tilde{\boldsymbol{\theta}})\right\}\\
&= O_p(g^{-\frac{3}{4}})\\
&= o_p(1),
\end{aligned}
$$

where $\tilde{\boldsymbol{\theta}}$ is between t_k and $\boldsymbol{\theta}_0$. Similarly, for the second term in (2.5.1), we have

$$
\begin{aligned}
&\eta^{-2}g^{-1}m^{-1}\sum_{k=1}^{N}\mathrm{E}\left[|l_{R\sigma_e^2}(t_k)-l_{R\sigma_e^2}(\boldsymbol{\theta}_0)-\mathrm{E}\left\{l_{R\sigma_e^2}(t_k)-l_{R\sigma_e^2}(\boldsymbol{\theta}_0)\right\}|^2\right]\\
&\leq \eta^{-2}g^{-1}m^{-1}\sum_{k=1}^{N}\mathrm{E}\left\{|l_{R\sigma_e^2}(t_k)-l_{R\sigma_e^2}(\boldsymbol{\theta}_0)|^2\right\}\\
&= \eta^{-2}g^{-1}m^{-1}\sum_{k=1}^{N}\mathrm{E}\left[|\sum_{i=1}^{g}\sum_{j=1}^{m}\left\{l_{ijR\sigma_e^2}(t_k)-l_{ijR\sigma_e^2}(\boldsymbol{\theta}_0)\right\}|^2\right]\\
&\leq \eta^{-2}g^{-1}m^{-1}\sum_{k=1}^{N}\sum_{i=1}^{g}\sum_{j=1}^{m}\mathrm{E}\left\{|l_{ijR\sigma_e^2}(t_k)-l_{ijR\sigma_e^2}(\boldsymbol{\theta}_0)|^2\right\}\\
&= \eta^{-2}g^{-1}m^{-1}\sum_{k=1}^{N}\sum_{i=1}^{g}\sum_{j=1}^{m}\mathrm{E}\left\{|\frac{\partial l_{ijR\sigma_e^2}(\tilde{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}}(t_k-\boldsymbol{\theta}_0)|^2\right\}\\
&= \eta^{-2}g^{-1}m^{-1}\sum_{k=1}^{N}\sum_{i=1}^{g}\sum_{j=1}^{m}\mathrm{E}\,|l_{ijR\sigma_e^2\sigma_\alpha^2}(\tilde{\boldsymbol{\theta}})(t_{k1}-\dot{\sigma_\alpha}^2)|^2+\eta^{-2}g^{-1}m^{-1}\sum_{k=1}^{N}\sum_{i=1}^{g}\sum_{j=1}^{m}\mathrm{E}\,|l_{ijR\sigma_e^2\sigma_e^2}(\tilde{\boldsymbol{\theta}})(t_{k2}-\dot{\sigma_e}^2)|^2\\
&\leq \eta^{-2}g^{-2}m^{-1}L\sum_{k=1}^{N}\sum_{i=1}^{g}\sum_{j=1}^{m}\mathrm{E}\left\{l^2_{ijR\sigma_e^2\sigma_\alpha^2}(\tilde{\boldsymbol{\theta}})\right\}+\eta^{-2}g^{-2}m^{-1}L\sum_{k=1}^{N}\sum_{i=1}^{g}\sum_{j=1}^{m}\mathrm{E}\left\{l^2_{ijR\sigma_e^2\sigma_e^2}(\tilde{\boldsymbol{\theta}})\right\}\\
&= O_p(g^{-\frac{3}{4}})\\
&= o_p(1).
\end{aligned}
$$

Next, we show that the difference between taking the supermum over a fine grid of points and over $\mathcal{B}$ is small. Using a Taylor expansion, we obtain

$$
\begin{aligned}
&\max_{1\le k\le N}\sup_{\boldsymbol{\theta}\in\mathcal{C}(t_k)} |\mathbf{K}_R^{-\frac{1}{2}}\left[\psi_R(\boldsymbol{\theta})-\psi_R(t_k)-\mathrm{E}\left\{\psi_R(\boldsymbol{\theta})-\psi_R(t_k)\right\}\right]| \\
&\le \max_{1\le k\le N}\sup_{\boldsymbol{\theta}\in\mathcal{C}(t_k)} g^{-\frac{1}{2}}|l_{R\sigma_\alpha^2}(\boldsymbol{\theta})-l_{R\sigma_\alpha^2}(t_k)-\mathrm{E}\left\{l_{R\sigma_\alpha^2}(\boldsymbol{\theta})-l_{R\sigma_\alpha^2}(t_k)\right\}| \\
&\quad+\max_{1\le k\le N}\sup_{\boldsymbol{\theta}\in\mathcal{C}(t_k)} g^{-\frac{1}{2}}m^{-\frac{1}{2}}|l_{R\sigma_e^2}(\boldsymbol{\theta})-l_{R\sigma_e^2}(t_k)-\mathrm{E}\left\{l_{R\sigma_e^2}(\boldsymbol{\theta})-l_{R\sigma_e^2}(t_k)\right\}| \\
&= \max_{1\le k\le N}\sup_{\boldsymbol{\theta}\in\mathcal{C}(t_k)} g^{-\frac{1}{2}}|\sum_{i=1}^{g}\left[l_{iR\sigma_\alpha^2}(\boldsymbol{\theta})-l_{iR\sigma_\alpha^2}(t_k)-\mathrm{E}\left\{l_{iR\sigma_\alpha^2}(\boldsymbol{\theta})-l_{iR\sigma_\alpha^2}(t_k)\right\}\right]| \\
&\quad+\max_{1\le k\le N}\sup_{\boldsymbol{\theta}\in\mathcal{C}(t_k)} g^{-\frac{1}{2}}m^{-\frac{1}{2}}|\sum_{i=1}^{g}\sum_{j=1}^{m}\left[l_{ijR\sigma_e^2}(\boldsymbol{\theta})-l_{ijR\sigma_e^2}(t_k)-\mathrm{E}\left\{l_{ijR\sigma_e^2}(\boldsymbol{\theta})-l_{ijR\sigma_e^2}(t_k)\right\}\right]| \\
&\le \max_{1\le k\le N}\sup_{\boldsymbol{\theta}\in\mathcal{C}(t_k)} g^{-\frac{1}{2}}\sum_{i=1}^{g}|l_{iR\sigma_\alpha^2}(\boldsymbol{\theta})-l_{iR\sigma_\alpha^2}(t_k)-\mathrm{E}\left\{l_{iR\sigma_\alpha^2}(\boldsymbol{\theta})-l_{iR\sigma_\alpha^2}(t_k)\right\}| \\
&\quad+\max_{1\le k\le N}\sup_{\boldsymbol{\theta}\in\mathcal{C}(t_k)} g^{-\frac{1}{2}}m^{-\frac{1}{2}}\sum_{i=1}^{g}\sum_{j=1}^{m}|l_{ijR\sigma_e^2}(\boldsymbol{\theta})-l_{ijR\sigma_e^2}(t_k)-\mathrm{E}\left\{l_{ijR\sigma_e^2}(\boldsymbol{\theta})-l_{ijR\sigma_e^2}(t_k)\right\}| \\
&\le \max_{1\le k\le N}\sup_{\boldsymbol{\theta}\in\mathcal{C}(t_k)} g^{-\frac{1}{2}}\sum_{i=1}^{g}|\left\{\frac{\partial l_{iR\sigma_\alpha^2}(\tilde{\boldsymbol{\theta}}_1)}{\partial\boldsymbol{\theta}}-\mathrm{E}\,\frac{\partial l_{iR\sigma_\alpha^2}(\boldsymbol{\theta}_1^*)}{\partial\boldsymbol{\theta}}\right\}(\boldsymbol{\theta}-t_k)| \\
&\quad+\max_{1\le k\le N}\sup_{\boldsymbol{\theta}\in\mathcal{C}(t_k)} g^{-\frac{1}{2}}m^{-\frac{1}{2}}\sum_{i=1}^{g}\sum_{j=1}^{m}|\left\{\frac{\partial l_{ijR\sigma_e^2}(\tilde{\boldsymbol{\theta}}_2)}{\partial\boldsymbol{\theta}}-\mathrm{E}\,\frac{\partial l_{ijR\sigma_e^2}(\boldsymbol{\theta}_2^*)}{\partial\boldsymbol{\theta}}\right\}(\boldsymbol{\theta}-t_k)| \\
&\le \max_{1\le k\le N}\sup_{\boldsymbol{\theta}\in\mathcal{C}(t_k)} g^{-\frac{1}{2}}\sum_{i=1}^{g}\left[|\left\{l_{iR\sigma_\alpha^2\sigma_\alpha^2}(\tilde{\boldsymbol{\theta}}_1)-\mathrm{E}\,l_{iR\sigma_\alpha^2\sigma_\alpha^2}(\boldsymbol{\theta}_1^*)\right\}(\dot{\sigma_\alpha}^2-t_{k1})|\right. \\
&\quad\left.+|\left\{l_{iR\sigma_\alpha^2\sigma_e^2}(\tilde{\boldsymbol{\theta}}_1)-\mathrm{E}\,l_{iR\sigma_\alpha^2\sigma_e^2}(\boldsymbol{\theta}_1^*)\right\}(\dot{\sigma_e}^2-t_{k2})|\right] \\
&\quad+\max_{1\le k\le N}\sup_{\boldsymbol{\theta}\in\mathcal{C}(t_k)} g^{-\frac{1}{2}}m^{-\frac{1}{2}}\sum_{i=1}^{g}\sum_{j=1}^{m}\left[|\left\{l_{ijR\sigma_e^2\sigma_\alpha^2}(\tilde{\boldsymbol{\theta}}_2)-\mathrm{E}\,l_{ijR\sigma_e^2\sigma_\alpha^2}(\boldsymbol{\theta}_2^*)\right\}(\dot{\sigma_\alpha}^2-t_{k1})|\right. \\
&\quad\left.+|\left\{l_{ijR\sigma_e^2\sigma_e^2}(\tilde{\boldsymbol{\theta}}_1)-\mathrm{E}\,l_{ijR\sigma_e^2\sigma_e^2}(\boldsymbol{\theta}_1^*)\right\}(\dot{\sigma_e}^2-t_{k2})|\right]
\end{aligned}
$$

where $\tilde{\boldsymbol{\theta}}_1, \tilde{\boldsymbol{\theta}}_2, \boldsymbol{\theta}_1^*$ and $\boldsymbol{\theta}_2^*$ are between t_k and $\boldsymbol{\theta}$. Since $\boldsymbol{\theta}\in\mathcal{C}(t_k)$, then we have $|(\dot{\sigma_\alpha}^2-t_{k1})|\le g^{-\frac{3}{4}}M$ and $|(\dot{\sigma_e}-t_{k2})|\le g^{-\frac{3}{4}}m^{-\frac{1}{2}}M$. From (2.4.22) and (2.4.25), we can find that $|l_{iR\sigma_\alpha^2\sigma_\alpha^2}(\tilde{\boldsymbol{\theta}}_1)-\mathrm{E}\,l_{iR\sigma_\alpha^2\sigma_\alpha^2}(\boldsymbol{\theta}_1^*)|$, $|l_{iR\sigma_\alpha^2\sigma_e^2}(\tilde{\boldsymbol{\theta}}_1)-\mathrm{E}\,l_{iR\sigma_\alpha^2\sigma_e^2}(\boldsymbol{\theta}_1^*)|$, $|l_{ijR\sigma_e^2\sigma_\alpha^2}(\tilde{\boldsymbol{\theta}}_2)-\mathrm{E}\,l_{ijR\sigma_e^2\sigma_\alpha^2}(\boldsymbol{\theta}_2^*)|$, and $|l_{ijR\sigma_e^2\sigma_e^2}(\tilde{\boldsymbol{\theta}}_1)-\mathrm{E}\,l_{ijR\sigma_e^2\sigma_e^2}(\boldsymbol{\theta}_1^*)|$ are all bounded. Thus, it follows that

$$
\begin{aligned}
&\max_{1\le k\le N}\sup_{\boldsymbol{\theta}\in\mathcal{C}(t_k)} g^{-\frac{1}{2}}\sum_{i=1}^{g}\left[|\left\{l_{iR\sigma_\alpha^2\sigma_\alpha^2}(\tilde{\boldsymbol{\theta}}_1)-\mathrm{E}\,l_{iR\sigma_\alpha^2\sigma_\alpha^2}(\boldsymbol{\theta}_1^*)\right\}(\dot{\sigma_\alpha}^2-t_{k1})|\right. \\
&\quad\left.+|\left\{l_{iR\sigma_\alpha^2\sigma_e^2}(\tilde{\boldsymbol{\theta}}_1)-\mathrm{E}\,l_{iR\sigma_\alpha^2\sigma_e^2}(\boldsymbol{\theta}_1^*)\right\}(\dot{\sigma_e}^2-t_{k2})|\right] \\
&\le \max_{1\le k\le N}\sup_{\boldsymbol{\theta}\in\mathcal{C}(t_k)} g^{-1-\frac{1}{4}}L\sum_{i=1}^{g}\left\{|l_{iR\sigma_\alpha^2\sigma_\alpha^2}(\tilde{\boldsymbol{\theta}}_1)-\mathrm{E}\,l_{iR\sigma_\alpha^2\sigma_\alpha^2}(\boldsymbol{\theta}_1^*)|+m^{-\frac{1}{2}}|l_{iR\sigma_\alpha^2\sigma_e^2}(\tilde{\boldsymbol{\theta}}_1)-\mathrm{E}\,l_{iR\sigma_\alpha^2\sigma_e^2}(\boldsymbol{\theta}_1^*)|\right\} \\
&= O_p(g^{-\frac{1}{4}})
\end{aligned}
$$

and

$$\max_{1\le k\le N}\sup_{\boldsymbol\theta\in\mathcal{C}(t_k)} g^{-\frac{1}{2}}m^{-\frac{1}{2}}\sum_{i=1}^{g}\sum_{j=1}^{m}\Big[|\Big\{l_{ijR\sigma_e^2\sigma_\alpha^2}(\tilde{\boldsymbol\theta}_2)-\mathrm{E}\,l_{ijR\sigma_e^2\sigma_\alpha^2}(\boldsymbol\theta_2^*)\Big\}(\dot{\sigma_\alpha}^2-t_{k1})|$$
$$+|\Big\{l_{ijR\sigma_e^2\sigma_e^2}(\tilde{\boldsymbol\theta}_1)-\mathrm{E}\,l_{ijR\sigma_e^2\sigma_e^2}(\boldsymbol\theta_1^*)\Big\}(\dot{\sigma_e}^2-t_{k2})|\Big]$$
$$\le\max_{1\le k\le N}\sup_{\boldsymbol\theta\in\mathcal{C}(t_k)} g^{-1-\frac{1}{4}}m^{-\frac{1}{2}}L\sum_{i=1}^{g}\sum_{j=1}^{m}|l_{ijR\sigma_e^2\sigma_\alpha^2}(\tilde{\boldsymbol\theta}_2)-\mathrm{E}\,l_{ijR\sigma_e^2\sigma_\alpha^2}(\boldsymbol\theta_2^*)|$$
$$+g^{-1-\frac{1}{4}}L\sum_{i=1}^{g}\sum_{j=1}^{m}\max_{1\le k\le N}\sup_{\boldsymbol\theta\in\mathcal{C}(t_k)}|\Big\{l_{ijR\sigma_e^2\sigma_e^2}(\tilde{\boldsymbol\theta}_1)-\mathrm{E}\,l_{ijR\sigma_e^2\sigma_e^2}(\boldsymbol\theta_1^*)\Big\}(\dot{\sigma_e}^2-t_{k2})|$$
$$=o_p(1).$$

$\square$

Lemma 11. *Suppose Condition A holds. Then, as $g,m\to\infty$,*

$$\sup_{|\mathbf{K}_R^{\frac{1}{2}}(\boldsymbol\theta-\boldsymbol\theta_0)|\le M}|\mathbf{K}_R^{-\frac{1}{2}}\left[\mathrm{E}\{\psi_R(\boldsymbol\theta)-\psi_R(\boldsymbol\theta_0)\}-\mathrm{E}\{\nabla\psi_R(\boldsymbol\theta_0)\}(\boldsymbol\theta-\boldsymbol\theta_0)\right]|=o_p(1).$$

Proof. Write

$$\sup_{|\mathbf{K}_R^{\frac{1}{2}}(\boldsymbol\theta-\boldsymbol\theta_0)|\le M}|\mathbf{K}_R^{-\frac{1}{2}}\left[\mathrm{E}\{\psi_R(\boldsymbol\theta)-\psi_R(\boldsymbol\theta_0)\}-\mathrm{E}\{\nabla\psi_R(\boldsymbol\theta_0)\}(\boldsymbol\theta-\boldsymbol\theta_0)\right]|$$
$$\le\sup_{|\mathbf{K}_R^{\frac{1}{2}}(\boldsymbol\theta-\boldsymbol\theta_0)|\le M}|g^{-\frac{1}{2}}\left[\mathrm{E}\left\{l_{R\sigma_\alpha^2}(\boldsymbol\theta)-l_{R\sigma_\alpha^2}(\boldsymbol\theta_0)\right\}-\mathrm{E}\left\{l_{R\sigma_\alpha^2\sigma_\alpha^2}(\boldsymbol\theta_0)(\sigma_\alpha^2-\dot{\sigma_\alpha}^2)+l_{R\sigma_\alpha^2\sigma_e^2}(\boldsymbol\theta_0)(\sigma_e^2-\dot{\sigma_e}^2)\right\}\right]|$$
$$+\sup_{|\mathbf{K}_R^{\frac{1}{2}}(\boldsymbol\theta-\boldsymbol\theta_0)|\le M}|g^{-\frac{1}{2}}m^{-\frac{1}{2}}\left[\mathrm{E}\left\{l_{R\sigma_e^2}(\boldsymbol\theta)-l_{R\sigma_e^2}(\boldsymbol\theta_0)\right\}-\mathrm{E}\left\{l_{R\sigma_e^2\sigma_\alpha^2}(\boldsymbol\theta_0)(\sigma_\alpha^2-\dot{\sigma_\alpha}^2)+l_{R\sigma_e^2\sigma_e^2}(\boldsymbol\theta_0)(\sigma_e^2-\dot{\sigma_e}^2)\right\}\right]|$$
$$\le\sup_{|\mathbf{K}_R^{\frac{1}{2}}(\boldsymbol\theta-\boldsymbol\theta_0)|\le M}g^{-\frac{1}{2}}\sum_{i=1}^{g}|\,\mathrm{E}\left\{l_{iR\sigma_\alpha^2}(\boldsymbol\theta)-l_{iR\sigma_\alpha^2}(\boldsymbol\theta_0)\right\}-\mathrm{E}\left\{l_{iR\sigma_\alpha^2\sigma_\alpha^2}(\boldsymbol\theta_0)(\sigma_\alpha^2-\dot{\sigma_\alpha}^2)+l_{iR\sigma_\alpha^2\sigma_e^2}(\boldsymbol\theta_0)(\sigma_e^2-\dot{\sigma_e}^2)\right\}|$$
$$+\sup_{|\mathbf{K}_R^{\frac{1}{2}}(\boldsymbol\theta-\boldsymbol\theta_0)|\le M}g^{-\frac{1}{2}}m^{-\frac{1}{2}}\sum_{i=1}^{g}\sum_{j=1}^{m}|\,\mathrm{E}\left\{l_{ijR\sigma_e^2}(\boldsymbol\theta)-l_{ijR\sigma_e^2}(\boldsymbol\theta_0)\right\}$$
$$-\mathrm{E}\left\{l_{ijR\sigma_e^2\sigma_\alpha^2}(\boldsymbol\theta_0)(\sigma_\alpha^2-\dot{\sigma_\alpha}^2)+l_{ijR\sigma_e^2\sigma_e^2}(\boldsymbol\theta_0)(\sigma_e^2-\dot{\sigma_e}^2)\right\}|$$
$$\le\sup_{|\mathbf{K}_R^{\frac{1}{2}}(\boldsymbol\theta-\boldsymbol\theta_0)|\le M}g^{-\frac{1}{2}}\sum_{i=1}^{g}|\,\mathrm{E}\Big\{l_{iR\sigma_\alpha^2\sigma_\alpha^2}(\tilde{\boldsymbol\theta})-l_{iR\sigma_\alpha^2\sigma_\alpha^2}(\boldsymbol\theta_0)\Big\}(\sigma_\alpha^2-\dot{\sigma_\alpha}^2)+\mathrm{E}\Big\{l_{iR\sigma_\alpha^2\sigma_e^2}(\tilde{\boldsymbol\theta})-l_{iR\sigma_\alpha^2\sigma_e^2}(\boldsymbol\theta_0)\Big\}(\sigma_e^2-\dot{\sigma_e}^2)|$$
$$+\sup_{|\mathbf{K}_R^{\frac{1}{2}}(\boldsymbol\theta-\boldsymbol\theta_0)|\le M}g^{-\frac{1}{2}}m^{-\frac{1}{2}}\sum_{i=1}^{g}\sum_{j=1}^{m}|\,\mathrm{E}\left\{l_{ijR\sigma_e^2\sigma_\alpha^2}(\boldsymbol\theta^*)-l_{ijR\sigma_e^2\sigma_\alpha^2}(\boldsymbol\theta_0)\right\}(\sigma_\alpha^2-\dot{\sigma_\alpha}^2)$$
$$+\mathrm{E}\left\{l_{ijR\sigma_e^2\sigma_e^2}(\boldsymbol\theta^*)-l_{ijR\sigma_e^2\sigma_e^2}(\boldsymbol\theta_0)\right\}(\sigma_e^2-\dot{\sigma_e}^2)|,$$

where both $\tilde{\boldsymbol\theta}$ and $\boldsymbol\theta^*$ are between $\boldsymbol\theta$ and $\boldsymbol\theta_0$. From (2.4.25) and (2.4.26), we can find that when $|\mathbf{K}_R^{\frac{1}{2}}(\boldsymbol\theta-\boldsymbol\theta_0)|\le M$, and as $g,m\to\infty$, we can obtain that $\mathrm{E}\{l_{iR\sigma_\alpha^2\sigma_\alpha^2}(\tilde{\boldsymbol\theta})-l_{iR\sigma_\alpha^2\sigma_\alpha^2}(\boldsymbol\theta_0)\}=$

$O(m^{-1})$, $\mathrm{E}\{l_{iR\sigma_\alpha^2\sigma_e^2}(\tilde{\boldsymbol{\theta}}) - l_{iR\sigma_\alpha^2\sigma_e^2}(\boldsymbol{\theta}_0)\} = O(m^{-1})$, $\mathrm{E}\{l_{ijR\sigma_e^2\sigma_\alpha^2}(\tilde{\boldsymbol{\theta}}) - l_{ijR\sigma_e^2\sigma_\alpha^2}(\boldsymbol{\theta}_0)\} = O(m^{-1})$ and $\mathrm{E}\{l_{ijR\sigma_e^2\sigma_e^2}(\tilde{\boldsymbol{\theta}}) - l_{ijR\sigma_e^2\sigma_e^2}(\boldsymbol{\theta}_0)\} = O(m^{-1})$. Thus the required result follows lemma 11. □

2.5.1 Proof of the main theorem

Proof. We write

$$\begin{aligned}\mathbf{K}_R^{-\frac{1}{2}}\psi(\boldsymbol{\theta}) = \mathbf{K}_R^{-\frac{1}{2}}\left[\psi_R(\boldsymbol{\theta}) - \psi_R(\boldsymbol{\theta}_0) - \mathrm{E}\left\{\psi_R(\boldsymbol{\theta}) - \psi_R(\boldsymbol{\theta}_0)\right\}\right] + \mathbf{K}_R^{-\frac{1}{2}}\left[\psi_R(\boldsymbol{\theta}_0) - \xi_R(\boldsymbol{\theta}_0)\right] \\ + \mathbf{K}_R^{-\frac{1}{2}}\left[\mathrm{E}\left\{\psi_R(\boldsymbol{\theta}) - \psi_R(\boldsymbol{\theta}_0) - \nabla\psi_R(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\right\}\right] \\ + \mathbf{K}_R^{-\frac{1}{2}}\left[\xi_R(\boldsymbol{\theta}_0) + \mathrm{E}\left\{\nabla\psi_R(\boldsymbol{\theta}_0)\right\}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\right].\end{aligned}$$

From (2.4.18), we have

$$\mathbf{K}_R^{-\frac{1}{2}}\left\{\psi_R(\boldsymbol{\theta}_0) - \xi_R(\boldsymbol{\theta}_0)\right\} = o_p(1).$$

Moreover, from lemma 10 and lemma 11, we have

$$\sup_{|\mathbf{K}_R^{\frac{1}{2}}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)|\leq M} \left|\mathbf{K}_R^{-\frac{1}{2}}\left[\psi_R(\boldsymbol{\theta}) - \psi_R(\boldsymbol{\theta}_0) - \mathrm{E}\left\{\psi_R(\boldsymbol{\theta}) - \psi_R(\boldsymbol{\theta}_0)\right\}\right]\right| = o_p(1) \qquad \text{and}$$

$$\sup_{|\mathbf{K}_R^{\frac{1}{2}}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)|\leq M} \left|\mathbf{K}_R^{-\frac{1}{2}}\left[\mathrm{E}\left\{\psi_R(\boldsymbol{\theta}) - \psi_R(\boldsymbol{\theta}_0)\right\} - \mathrm{E}\left\{\nabla\psi_R(\boldsymbol{\theta}_0)\right\}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\right]\right| = o_p(1).$$

Thus, on combining the above equations, we can see that uniformly on $\mathbf{K}_R^{\frac{1}{2}}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \leq M$,

$$\mathbf{K}_R^{-\frac{1}{2}}\psi_R(\boldsymbol{\theta}) = \mathbf{K}_R^{-\frac{1}{2}}\xi_R(\boldsymbol{\theta}_0)\mathbf{B}_{Rn}\mathbf{K}_R^{\frac{1}{2}}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + o_p(1), \tag{2.5.2}$$

where $\mathbf{B}_{Rn} = -\mathbf{K}_R^{-\frac{1}{2}}\,\mathrm{E}\,\nabla\psi_R(\boldsymbol{\theta}_0)\mathbf{K}_R^{-\frac{1}{2}}$. The right-hand side of (2.5.2) is mainly determined by the negative of the quadratic form of $\mathbf{B}_{Rn}$ and is negative for M sufficiently large. Therefore, applying to Result 6.3.4 of Ortega and Rheinboldt (1973), we have that a solution to the estimating equations exists in probability and satisfies $|\mathbf{K}_R^{-\frac{1}{2}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)| = O_p(1)$. The asymptotic representation follows from lemma 10 and lemma 11. Lemma 9 implies that the Lyapounov condition holds and hence the asymptotic normality follows. □

Chapter 3 Unbalanced Data: Asymptotic Analysis of ML and REML Estimates for Hierarchical Linear Mixed Models with Large Cluster Sizes

Part of this chapter was submitted as a paper to a statistical journal for publication.

3.1 Introduction

In this chapter, we study asymptotic properties of ML and REML estimates for nested error regression model in more general case, unbalanced data setting, which is necessarily more complicated for asymptotic derivation and computing. Since in summing the squares of the deviations of the group means from the grand mean, each group has been given a different weight m_i i.e.,the number of observations in group i, we can not use Kronecker products to simplify derivation in many formulas and equations, especially in studying asymptotic properties of ML and REML estimates. Compared with the REML approach based on the balanced data case in Chapter 2, we first construct adjusted maximum likelihood (AML) estimates, then obtain asymptotic properties of the REML estimators, as the REML estimating equation is the profile version of AML estimating equation.

Suppose that (y_{ij}, x_{ij}) is the jth observation, $j = 1, \ldots, m_i$, from the ith cluster, $i = 1, \ldots, g$, where y_{ij} is a scalar response variable, x_{ij} is a scalar explanatory variable or covariate and $m_L \leq m_i \leq m_U$. As previously in Section 2.1, the nested error regression model specifies that

$$y_{ij} = \mu(x_{ij}) + \alpha_i + e_{ij}, \qquad j = 1, \ldots, m_i,\ i = 1, \ldots, g, \tag{3.1.1}$$

where $\mu(x_{ij})$ is the conditional mean of the response given the covariate x_{ij}, α_i is a random effect representing a random intercept or cluster effect, and e_{ij} is an error term. We assume that the $\{\alpha_i\}$ and $\{e_{ij}\}$ are mutually independent with mean zero and variances (called the variance components) σ^2_α and σ^2_e, respectively. We can estimate both variance components consistently when

$g \to \infty$ with m_U bounded but we need to let $m_L \to \infty$ in order to estimate the random effects $\{\alpha_i\}$ consistently. If $m_L \to \infty$ but g is held fixed, then we can estimate the within cluster variance σ_e^2 consistently but not the between cluster variance σ_α^2. These settings motivate considering both $g \to \infty$ and $m_L \to \infty$.

We consider the same framework from Chapter 2 by setting

$$\mu(x_{ij}) = \begin{cases} \beta_0 + \beta_1 x_i & \text{if } x_{ij} = x_i \\ \beta_0 + \beta_1 \bar{x}_{i.} + \beta_2(x_{ij} - \bar{x}_{i.}) & \text{otherwise} \end{cases}, \tag{3.1.2}$$

where the βs are fixed effect parameters, called regression parameters or fixed effects, and $\bar{x}_{i.} = m_i^{-1} \sum_{j=1}^{m_i} x_{ij}$ is the within cluster mean of the explanatory variable. In (3.1.2) we distinguish a between cluster covariate that is constant within clusters ($x_{ij} = x_i$) from a within cluster covariate that varies within clusters. This distinction is often ignored by using $\mu(x_{ij}) = \beta_0 + \beta_1 x_{ij}$ for both cases.

We extend the finding of Chapter 2 to the unbalance case and show that $\mu(x_{ij}) = \beta_0 + \beta_1 x_{ij}$ does not work asymptotically (when $m_L \to \infty$) for a within cluster covariate. The reason is that the information for within cluster parameters $\boldsymbol{\omega}^{(w)} = [\beta_2, \sigma_e^2]^T$ grows with $n = \sum_{i=1}^{g} m_i$, while the information for between cluster parameters $\boldsymbol{\omega}^{(b)} = [\beta_0, \beta_1, \sigma_\alpha^2]^T$ grows with g. The information for the coefficient of a within cluster covariate that has non-zero mean grows with both n and g, thereby making it difficult to normalize to control both sources of information at the same time. In contrast, if we partition the within cluster covariate into a between cluster component and a pure within cluster component (which has all within cluster means zero), as in (3.1.2), we obtain separate coefficients for which the information grows with n and g separately, and normalization becomes more straightforward.

Under the assumption that the random effects and errors are normally distributed, the likelihood for the parameters $\boldsymbol{\omega} = [\boldsymbol{\omega}^{(b)T}, \boldsymbol{\omega}^{(w)T}]^T$ and the REML criterion (an adjusted profile likelihood for $\boldsymbol{\theta} = [\sigma_\alpha^2, \sigma_e^2]^T$ after eliminating $\boldsymbol{\beta} = [\beta_0, \beta_1, \beta_2]$) can be obtained explicitly. We refer to these functions as the likelihood and the REML criterion, selectively for (3.1.1) and the values $\hat{\boldsymbol{\omega}}$ and $\hat{\boldsymbol{\theta}}$ that maximize them as the maximum likelihood estimator and the REML estimator, respectively regardless of whether the normality assumption holds or not. The asymptotic distributions of the maximum likelihood and REML estimators have been obtained under a variety of conditions by various authors (see the review in Section 2.1).

We rewrite (3.1.1) in matrix from as

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{1}_{m_i} \alpha_i + \mathbf{e}_i, \tag{3.1.3}$$

$\mathbf{y}_i = (y_{i1}, \ldots, y_{im})^T$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^T$, $\mathbf{X}_i = [\mathbf{1}_{m_i}, \mathbf{1}_{m_i}\bar{x}_{i.}, \mathbf{x}_i - \mathbf{1}_{m_i}\bar{x}_{i.}]$ with $\mathbf{1}_m$ is the m_i-vector of ones and $\mathbf{e}_i = (e_{i1}, \ldots, e_{im_i})^T$. Let $\mathbf{y}^T = (\mathbf{y}_1^T, \ldots, \mathbf{y}_g^T)$ and $\mathbf{x}^T = (\mathbf{x}_1^T, \ldots, \mathbf{x}_g^T)$. Under this model, we have

$$\mathrm{E}(\mathbf{y}_i) = \mathbf{1}_{m_i}\beta_0 + \mathbf{1}_{m_i}\bar{x}_{i.}\beta_1 + (\mathbf{x}_i - \mathbf{1}_{m_i}\bar{x}_{i.})\beta_2, \qquad \mathrm{Var}(\mathbf{y}_i) = \sigma_\alpha^2\mathbf{J}_{m_i} + \sigma_e^2\mathbf{I}_{m_i},$$

$$\text{and} \qquad \mathrm{Var}(\mathbf{y}) = \mathrm{diag}[\sigma_\alpha^2\mathbf{J}_{m_1} + \sigma_e^2\mathbf{I}_{m_1}, \ldots, \sigma_\alpha^2\mathbf{J}_{m_g} + \sigma_e^2\mathbf{I}_{m_g}].$$

We find that the above equations are more general in structure i.e., no Kronecker Products, compared to the balanced data case in Chapter 2, thus complicating algebraic developments. Regardless, we proceed to partition the parameters in (3.1.3) into within cluster and between cluster parameters. Let $\boldsymbol{\omega}^{(b)} = (\beta_0, \beta_1, \sigma_\alpha^2)^T$ denote the between cluster parameters, and let $\boldsymbol{\omega}^{(w)} = (\beta_2, \sigma_e^2)^T$ denote the within cluster parameters. Let $\boldsymbol{\omega} = \mathrm{vec}(\boldsymbol{\omega}^{(b)}, \boldsymbol{\omega}^{(w)}) = (\beta_0, \beta_1, \sigma_\alpha^2, \beta_2, \sigma_e^2)^T$ and let the true parameter vector be $\boldsymbol{\omega}_0 = (\dot{\beta}_0, \dot{\beta}_2, \dot{\sigma_\alpha}^2, \dot{\beta}_1, \dot{\sigma_e}^2)^T$. Let $\boldsymbol{\theta} = (\sigma_\alpha^2, \sigma_e^2)^T$ and $\boldsymbol{\theta}_0 = (\dot{\sigma_\alpha}^2, \dot{\sigma_e}^2)^T$, such that we can write

$$\mathbf{V}_i^{-1}(\boldsymbol{\theta}) = \frac{1}{\sigma_e^2}\mathbf{I}_{m_i} - \frac{\sigma_\alpha^2}{\sigma_e^2(\sigma_e^2 + m_i\sigma_\alpha^2)}\mathbf{J}_{m_i} \qquad \text{and} \qquad \|\mathbf{V}(\boldsymbol{\theta})\| = \prod_{i=1}^{g} \sigma_e^{2(m_i-1)}(\sigma_e^2 + m_i\sigma_\alpha^2)$$

For easy calculation, we denote

$$\bar{y}_{i.} = \frac{1}{m_i}\sum_{j=1}^{m_i} y_{ij}, \qquad \bar{y}_{..} = \frac{1}{n}\sum_{i=1}^{g}\sum_{j=1}^{m_i} y_{ij}, \qquad \bar{x}_{..} = \frac{1}{n}\sum_{i=1}^{g}\sum_{j=1}^{m_i} x_{ij},$$

$$S_{ijw}^{xy} = (x_{ij} - \bar{x}_{i.})(y_{ij} - \bar{y}_{i.}), \qquad \text{and} \qquad S_w^{xy} = \sum_{i=1}^{g}\sum_{j=1}^{m_i} S_{ijw}^{xy}.$$

To simplify notation, let $S_w^{xx} = S_w^x$, and $S_w^{yy} = S_w^y$. When the random effects α_i and e_{ij} in (3.1.1) are all normally distributed, it follows that the observations $\mathbf{y}_i$ are also normally distributed. In turn we can write the density of $\mathbf{y}$ given $\mathbf{x}$ as

$$L(\mathbf{y}; \boldsymbol{\omega}) = (2\pi)^{-\frac{n}{2}}|\mathbf{V}(\boldsymbol{\theta})|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{g}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^T\mathbf{V}_i^{-1}(\boldsymbol{\theta})(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})\right\}.$$

Furthermore, we have

$$
\begin{aligned}
&\sum_{i=1}^{g}(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})^T\mathbf{V}_i^{-1}(\boldsymbol{\theta}_0)(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})\\
&=\sum_{i=1}^{g}(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})^T\left\{\frac{1}{\sigma_e^2}\mathbf{I}_{m_i}-\frac{\sigma_\alpha^2}{\sigma_e^2(\sigma_e^2+m_i\sigma_\alpha^2)}\mathbf{J}_{m_i}\right\}(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})\\
&=\frac{1}{\sigma_e^2}\sum_{i=1}^{g}\{\mathbf{y}_i-\mathbf{1}_{m_i}\beta_0-\mathbf{1}_{m_i}\bar{x}_{i.}\beta_1-(\mathbf{x}_i-\mathbf{1}_{m_i}\bar{x}_{i.})\beta_2\}^T\{\mathbf{y}_i-\mathbf{1}_{m_i}\beta_0-\mathbf{1}_{m_i}\bar{x}_{i.}\beta_1-(\mathbf{x}_i-\mathbf{1}_{m_i}\bar{x}_{i.})\beta_2\}\\
&\qquad-\sum_{i=1}^{g}\frac{\sigma_\alpha^2}{\sigma_e^2(\sigma_e^2+m_i\sigma_\alpha^2)}\left[\sum_{j=1}^{m_i}\{y_{ij}-\beta_0-\bar{x}_{i.}\beta_1-(x_{ij}-\bar{x}_{i.})\beta_2\}\right]^2\\
&=\frac{1}{\sigma_e^2}\sum_{i=1}^{g}\sum_{j=1}^{m_i}(y_{ij}-\beta_0-\bar{x}_{i.}\beta_1-x_{ij}\beta_2+\bar{x}_{i.}\beta_2)^2-\sum_{i=1}^{g}\frac{\sigma_\alpha^2}{\sigma_e^2(\sigma_e^2+m_i\sigma_\alpha^2)}\left\{\sum_{j=1}^{m_i}(y_{ij}-\beta_0-\bar{x}_{i.}\beta_1-x_{ij}\beta_2+\bar{x}_{i.}\beta_2)\right\}^2\\
&=\frac{1}{\sigma_e^2}\sum_{i=1}^{g}\sum_{j=1}^{m_i}(y_{ij}-\beta_0-\bar{x}_{i.}\beta_1-x_{ij}\beta_2+\bar{x}_{i.}\beta_2)^2-\sum_{i=1}^{g}\frac{{m_i}^2\sigma_\alpha^2}{\sigma_e^2(\sigma_e^2+m_i\sigma_\alpha^2)}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1-\bar{x}_{i.}\beta_2+\bar{x}_{i.}\beta_2)^2\\
&=\frac{1}{\sigma_e^2}\sum_{i=1}^{g}\sum_{j=1}^{m_i}(y_{ij}-\beta_0-\bar{x}_{i.}\beta_1-x_{ij}\beta_2+\bar{x}_{i.}\beta_2)^2-\sum_{i=1}^{g}\frac{m_i}{\sigma_e^2}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1)^2+\sum_{i=1}^{g}\frac{m_i}{\sigma_e^2}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1)^2\\
&\qquad-\sum_{i=1}^{g}\frac{{m_i}^2\sigma_\alpha^2}{\sigma_e^2(\sigma_e^2+m_i\sigma_\alpha^2)}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1)^2\\
&=\frac{1}{\sigma_e^2}\sum_{i=1}^{g}\{(\mathbf{y}_i-\mathbf{1}_{m_i}\bar{y}_{i.})-(\mathbf{x}_i-\mathbf{1}_{m_i}\bar{x}_{i.})\beta_2\}^T\{(\mathbf{y}_i-\mathbf{1}_{m_i}\bar{y}_{i.})-(\mathbf{x}_i-\mathbf{1}_{m_i}\bar{x}_{i.})\beta_2\}+\sum_{i=1}^{g}\frac{m_i}{\sigma_e^2}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1)^2\\
&\qquad-\sum_{i=1}^{g}\frac{{m_i}^2\sigma_\alpha^2}{\sigma_e^2(\sigma_e^2+m_i\sigma_\alpha^2)}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1)^2\\
&=\frac{1}{\sigma_e^2}\sum_{i=1}^{g}\{(\mathbf{y}_i-\mathbf{1}_{m_i}\bar{y}_{i.})-(\mathbf{x}_i-\mathbf{1}_{m_i}\bar{x}_{i.})\beta_2\}^T\{(\mathbf{y}_i-\mathbf{1}_{m_i}\bar{y}_{i.})-(\mathbf{x}_i-\mathbf{1}_{m_i}\bar{x}_{i.})\beta_2\}\\
&\qquad+\sum_{i=1}^{g}\frac{m_i}{(\sigma_e^2+m_i\sigma_\alpha^2)}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1)^2\\
&=\frac{1}{\sigma_e^2}\sum_{i=1}^{g}\left\{(\mathbf{y}_i-\mathbf{1}_{m_i}\bar{y}_{i.})^T(\mathbf{y}_i-\mathbf{1}_{m_i}\bar{y}_{i.})-2(\mathbf{y}_i-\mathbf{1}_{m_i}\bar{y}_{i.})^T(\mathbf{x}_i-\mathbf{1}_{m_i}\bar{x}_{i.})\beta_2+(\mathbf{x}_i-\mathbf{1}_{m_i}\bar{x}_{i.})^T(\mathbf{x}_i-\mathbf{1}_{m_i}\bar{x}_{i.})\beta_2^2\right\}\\
&\qquad+\sum_{i=1}^{g}\frac{m_i}{(\sigma_e^2+m_i\sigma_\alpha^2)}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1)^2\\
&=\frac{1}{\sigma_e^2}(S_w^y-2S_w^{xy}\beta_2+S_w^x\beta_2^2)+\sum_{i=1}^{g}\frac{m_i}{(\sigma_e^2+m_i\sigma_\alpha^2)}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1)^2.
\end{aligned}
$$

From the above, we find that in summing the squares of the deviations of the group means from the grand mean, each group has been given a different weight m_i, and so we are unable to simplify the derivation further when compared with Chapter 2. We rewrite the density of function of $\mathbf{y}$ as

$$
\begin{aligned}
&L(\mathbf{y};\boldsymbol{\omega})\\
&=(2\pi)^{-\frac{n}{2}}|\mathbf{V}(\boldsymbol{\theta})|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}\left\{\frac{1}{\sigma_e^2}(S_w^y-2S_w^{xy}\beta_2+S_w^x\beta_2^2)+\sum_{i=1}^g\frac{m_i}{(\sigma_e^2+m_i\sigma_\alpha^2)}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1)^2\right\}\right].
\end{aligned}
$$

Moreover, the log-likelihood function is given by

$$
\begin{aligned}
l(\mathbf{y};\boldsymbol{\omega})=&-\frac{1}{2}n\log 2\pi-\frac{1}{2}\sum_{i=1}^g\log(\sigma_e^2+m_i\sigma_\alpha^2)-\frac{n-g}{2}\log\sigma_e^2-\frac{1}{2\sigma_e^2}(S_w^y-2S_w^{xy}\beta_2+S_w^x\beta_2^2)\\
&-\sum_{i=1}^g\frac{m_i}{2(\sigma_e^2+m_i\sigma_\alpha^2)}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1)^2.
\end{aligned}
\tag{3.1.4}
$$

To maximize $l(\mathbf{y};\boldsymbol{\omega})$, we differentiate (3.1.4) with respect to $\beta_0,\beta_1,\beta_2,\sigma_\alpha^2$ and σ_e^2, respectively. Let $\psi(\boldsymbol{\omega})^T=[\psi^{(b)}(\boldsymbol{\omega})^T,\psi^{(w)}(\boldsymbol{\omega})^T]$, where $\psi^{(b)}(\boldsymbol{\omega})^T=[l_{\beta_0}(\boldsymbol{\omega}),l_{\beta_1}(\boldsymbol{\omega}),l_{\sigma_\alpha^2}(\boldsymbol{\omega})]$ and $\psi^{(w)}(\boldsymbol{\omega})^T=[l_{\beta_2}(\boldsymbol{\omega}),l_{\sigma_e^2}(\boldsymbol{\omega})]$. Furthermore, we obtain the components of $\psi(\boldsymbol{\omega})$ as

$$
\begin{aligned}
l_{\beta_0}(\boldsymbol{\omega})&=\sum_{i=1}^g\frac{m_i}{(\sigma_e^2+m_i\sigma_\alpha^2)}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1)\\
l_{\beta_1}(\boldsymbol{\omega})&=\sum_{i=1}^g\frac{m_i\bar{x}_{i.}}{(\sigma_e^2+m_i\sigma_\alpha^2)}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1)\\
l_{\sigma_\alpha^2}(\boldsymbol{\omega})&=-\frac{1}{2}\sum_{i=1}^g\frac{m_i}{(\sigma_e^2+m_i\sigma_\alpha^2)}+\sum_{i=1}^g\frac{{m_i}^2}{2(\sigma_e^2+m_i\sigma_\alpha^2)^2}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1)^2\\
l_{\beta_2}(\boldsymbol{\omega})&=\frac{1}{\sigma_e^2}S_w^{xy}-\frac{1}{\sigma_e^2}S_w^x\beta_2\\
l_{\sigma_e^2}(\boldsymbol{\omega})&=-\sum_{i=1}^g\frac{1}{2(\sigma_e^2+m_i\sigma_\alpha^2)}-\frac{n-g}{2\sigma_e^2}+\frac{1}{2\sigma_e^4}(S_w^y-2S_w^{xy}\beta_2+S_w^x\beta_2^2)\\
&\quad+\sum_{i=1}^g\frac{m_i}{2(\sigma_e^2+m_i\sigma_\alpha^2)^2}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1)^2.
\end{aligned}
\tag{3.1.5}
$$

Compared with (2.1.5), we find that each equation in (3.1.5) has a different weight m_i . These weights will cause many troubles in later on, and motivated an alternative approach to deriving desired asymptotic results.

3.2 Conditions

To derive the asymptotic properties of $\hat{\boldsymbol{\omega}}$ in (3.1.5), we impose the following condition. Note that all the expectations are under the true parameter $\boldsymbol{\omega}_0$.

Condition B

1. The model (3.1.1) holds with true parameters $\boldsymbol{\omega}_0$ inside the parameter space Ω.

2. The number of clusters $g \to \infty$, and the minimum cluster size $m_L \to \infty$.

3. The random variables $\{\alpha_i\}$ and $\{e_{ij}\}$ are independent and identically distributed and there exists a $\delta > 0$ such that $\mathrm{E}(|\alpha_i|^{4+\delta}) < \infty$ and $\mathrm{E}(|e_{ij}|^{4+\delta}) < \infty$ for all $i = 1, \ldots, g$ and $j = 1, \ldots, m_i$.

4. For the sequence of within cluster means $\{\bar{x}_{i.}\}$, there exist finite constants c_1 and c_2 such that $\lim_{g\to\infty} g^{-1} \sum_{i=1}^{g} \bar{x}_{i.}^k = c_k$, $k = 1, 2$ and a $\delta > 0$ such that $\lim_{g\to\infty} g^{-1} \sum_{i=1}^{g} |\bar{x}_{i.}|^{2+\delta} < \infty$. Also for the sequence $\{S_w^x\}$, there exists a constant c_3 such that $\lim_{g\to\infty} \lim_{m_L\to\infty} n^{-1} S_w^x = c_3 > 0$ and a $\delta > 0$ such that $\lim_{g\to\infty} \lim_{m_L\to\infty} n^{-1} \sum_{i=1}^{g} \sum_{j=1}^{m_i} |x_{ij} - \bar{x}_{i.}|^{2+\delta} < \infty$.

The above are mild conditions which are often satisfied in practice. Conditions B3 and B4 ensure that limits ensure that the limits required for the existence and finiteness of the asymptotic variance of the estimating function exist, and that we can establish an Lyapounov central limit theorem for the estimating function. Condition B4 holds, for example, if the x_{ij} are realizations of clustered random variables that have finite $4 + \lambda$ moments. For a pure within cluster covariate, we replace Condition B4 by

4'. For the sequence $\{x_{ij}\}$, there exists a finite constant c_3 such that $\lim_{g\to\infty} \lim_{m_L\to\infty} n^{-1} \sum_{i=1}^{g} \sum_{j=1}^{m_i} x_{ij}^2 = c_3$, and a $\delta > 0$ such that $\lim_{g\to\infty} \lim_{m_L\to\infty} n^{-1} \sum_{i=1}^{g} \sum_{j=1}^{m_i} |x_{ij}|^{2+\delta} < \infty$.

For a between cluster covariate $\{x_i\}$, we replace Condition B4 by

4". For the sequence $\{x_i\}$, there exist finite constants c_1 and c_2 such that $\lim_{g\to\infty} g^{-1} \sum_{i=1}^{g} x_i^k = c_k$, $k = 1, 2$, and a $\delta > 0$ such that $\lim_{g\to\infty} g^{-1} \sum_{i=1}^{g} |x_i|^{2+\delta} < \infty$.

3.2.1 Approximating the ML estimating equations evaluated at $\boldsymbol{\omega}_0$

The final central limit theorem is strongly driven by the behavior of the estimating equation evaluated at $\boldsymbol{\omega}_0$. We therefore start by developing a simple approximation to the estimating equation evaluated at $\boldsymbol{\omega}_0$. First, we write $\bar{y}_{i.} = \dot{\beta}_0 + \dot{\beta}_1 \bar{x}_{i.} + \alpha_i + \bar{e}_{i.}$, so that

$$y_{ij} - \bar{y}_{i.} = (x_{ij} - \bar{x}_{i.})\dot{\beta}_2 + (e_{ij} - \bar{e}_{i.}) \qquad \text{and} \qquad \bar{y}_{i.} - \dot{\beta}_0 - \dot{\beta}_1 \bar{x}_{i.} = \alpha_i + \bar{e}_{i.}$$

Then $\mathrm{Var}(\alpha_i) = \dot{\sigma_\alpha}^2$, $\mathrm{Var}(\bar{e}_{i.}) = \frac{\dot{\sigma_e}^2}{m_i}$, and thus

$$\bar{y}_{i.} - \dot{\beta}_0 - \dot{\beta}_1 \bar{x}_{i.} = \alpha_i + o_p(1) \tag{3.2.1}$$

and

$$(\bar{y}_{i.} - \dot{\beta}_0 - \dot{\beta}_1 \bar{x}_{i.})^2 = \alpha_i^2 + o_p(1). \tag{3.2.2}$$

Next, we obtain approximations for S_w^{xy} and S_w^y, respectively. For S_w^{xy}, we have

$$\begin{aligned}
S_w^{xy} &= \sum_{i=1}^{g} \sum_{j=1}^{m_i} (x_{ij} - \bar{x}_{i.})(y_{ij} - \bar{y}_{i.}) \\
&= \sum_{i=1}^{g} \sum_{j=1}^{m_i} \{(x_{ij} - \bar{x}_{i.})^2 \dot{\beta}_2 + (x_{ij} - \bar{x}_{i.})(e_{ij} - \bar{e}_{i.})\} \\
&= \sum_{i=1}^{g} \sum_{j=1}^{m_i} \{(x_{ij} - \bar{x}_{i.})^2 \dot{\beta}_2 - (x_{ij} - \bar{x}_{i.}) e_{ij}\} \\
&= S_w^x \dot{\beta}_2 + \sum_{i=1}^{g} \sum_{j=1}^{m_i} (x_{ij} - \bar{x}_{i.}) e_{ij},
\end{aligned} \tag{3.2.3}$$

such that

$$\mathrm{E}(S_w^{xy}) = S_w^x \dot{\beta}_2, \tag{3.2.4}$$

and

$$\begin{aligned}
\mathrm{Var}(S_w^{xy}) &= \sum_{i=1}^{g} \sum_{j=1}^{m_i} (x_{ij} - \bar{x}_{i.})^2 \,\mathrm{Var}(e_{ij}) \\
&= \dot{\sigma_e}^2 S_w^x \\
&= O(n).
\end{aligned}$$

For S_w^y, we write

$$\begin{aligned}
S_w^y &= \sum_{i=1}^{g} \sum_{j=1}^{m_i} (y_{ij} - \bar{y}_{i.})^2 \\
&= \sum_{i=1}^{g} \sum_{j=1}^{m_i} \{(x_{ij} - \bar{x}_{i.}) \dot{\beta}_2 + (e_{ij} - \bar{e}_{i.})\}^2 \\
&= \sum_{i=1}^{g} \sum_{j=1}^{m_i} \{(x_{ij} - \bar{x}_{i.})^2 \dot{\beta}_2^2 + 2(x_{ij} - \bar{x}_{i.}) \dot{\beta}_2 (e_{ij} - \bar{e}_{i.}) + (e_{ij} - \bar{e}_{i.})^2\} \\
&= S_w^x \dot{\beta}_2^2 + 2\dot{\beta}_2 \sum_{i=1}^{g} \sum_{j=1}^{m_i} (x_{ij} - \bar{x}_{i.}) e_{ij} + \sum_{i=1}^{g} \sum_{j=1}^{m_i} (e_{ij}^2 - \dot{\sigma_e}^2) + n\dot{\sigma_e}^2 - \sum_{i=1}^{g} m_i \bar{e}_{i.}^2,
\end{aligned} \tag{3.2.5}$$

such that

$$\mathrm{E}(S_w^y) = S_w^x \dot{\beta}_2^2 + n\dot{\sigma_e}^2 - g\dot{\sigma_e}^2 = S_w^x \dot{\beta}_2^2 + (n-g)\dot{\sigma_e}^2. \tag{3.2.6}$$

Next, we can show that

$$\mathrm{Var}\left\{\dot\beta_2\sum_{i=1}^{g}\sum_{j=1}^{m_i}(x_{ij}-\bar x_{i.})e_{ij}\right\}=\dot\beta_2^2\sum_{i=1}^{g}\sum_{j=1}^{m_i}(x_{ij}-\bar x_{i.})^2\,\mathrm{Var}(e_{ij})=\dot\beta_2^2S_w^x\dot{\sigma_e}^2=O(n),$$

$$\mathrm{Var}(\sum_{i=1}^{g}\sum_{j=1}^{m_i}e_{ij}^2)=\sum_{i=1}^{g}\sum_{j=1}^{m_i}\mathrm{Var}(e_{ij}^2)=\sum_{i=1}^{g}\sum_{j=1}^{m_i}\left\{\mathrm{E}(e_{ij}^4)-\dot{\sigma_e}^4\right\}=n\left\{\mathrm{E}(e_{ij}^4)-\dot{\sigma_e}^4\right\}=O(n),$$

and moreover

$$\begin{aligned}
\mathrm{Var}(m_i\bar e_{i.}^2)&=\mathrm{Var}\left\{{m_i}^{-1}(m_i\bar e_{i.})^2\right\}\\
&={m_i}^{-2}\,\mathrm{Var}\left\{(\sum_{j=1}^{m_i}e_{ij})^2\right\}\\
&={m_i}^{-2}\,\mathrm{Var}(\sum_{j=1}^{m_i}\sum_{k=1}^{m_i}e_{ij}e_{ik})\\
&={m_i}^{-2}\left\{\mathrm{E}(\sum_{j=1}^{m_i}\sum_{k=1}^{m_i}e_{ij}e_{ik})^2-\mathrm{E}^2(\sum_{j=1}^{m_i}\sum_{k=1}^{m_i}e_{ij}e_{ik})\right\}\\
&={m_i}^{-2}\sum_{j=1}^{m_i}\sum_{k=1}^{m_i}\sum_{r=1}^{m_i}\sum_{r=1}^{m_i}\mathrm{E}(e_{ij}e_{ik}e_{ir}e_{is})-{m_i}^{-2}\left\{\sum_{j=1}^{m_i}\sum_{k=1}^{m_i}\mathrm{E}(e_{ij}e_{ik})\right\}^2\\
&={m_i}^{-2}\sum_{j=1}^{m_i}\mathrm{E}(e_{ij}^4)+(3-3{m_i}^{-1})\dot{\sigma_e}^4-\dot{\sigma_e}^4\\
&={m_i}^{-1}\,\mathrm{E}(e_{ij}^4)+(2-3{m_i}^{-1})\dot{\sigma_e}^4\\
&=O(1).
\end{aligned}$$

The key step here was a result given in Lemma 1 [Cramér (1999), p345] . Combining the above results, along with $\mathrm{Var}(S_w^y)=O(n)$ and we obtain

$$S_w^y=n\dot{\sigma_e}^2+S_w^x\dot\beta_2^2+2\dot\beta_2\sum_{i=1}^{g}\sum_{j=1}^{m_i}(x_{ij}-\bar x_{i.})e_{ij}+\sum_{i=1}^{g}\sum_{j=1}^{m_i}(e_{ij}^2-\dot{\sigma_e}^2)+o_p(n^{\frac12}).\qquad(3.2.7)$$

Furthermore, combining (3.2.3) and (3.2.7), we obtain

$$S_w^y-2S_w^{xy}\dot\beta_2+S_w^x\dot\beta_2^2=n\dot{\sigma_e}^2+\sum_{i=1}^{g}\sum_{j=1}^{m_i}(e_{ij}^2-\dot{\sigma_e}^2)+o_p(n^{\frac12}).\qquad(3.2.8)$$

Combining (3.2.1), (3.2.2) and(3.2.8), we obtain

$$\mathbf{K}^{-\frac12}\psi(\boldsymbol\omega_0)=\mathbf{K}^{-\frac12}\xi(\boldsymbol\theta_0)+o_p(1),\qquad(3.2.9)$$

where $\mathbf{K} = \mathrm{diag}[g^{-\frac{1}{2}}\mathbf{I}_{(b)}, n^{-\frac{1}{2}}\mathbf{I}_{(w)}]$ and $\xi(\boldsymbol{\theta}_0)$ has components,

$$\xi_{\beta_0}(\boldsymbol{\theta}_0) = \sum_{i=1}^{g} \frac{m_i}{(\dot{\sigma_e}^2 + m_i\dot{\sigma_\alpha}^2)}\alpha_i,$$

$$\xi_{\beta_1}(\boldsymbol{\theta}_0) = \sum_{i=1}^{g} \frac{m_i}{(\dot{\sigma_e}^2 + m_i\dot{\sigma_\alpha}^2)}\bar{x}_{i.}\alpha_i,$$

$$\xi_{\sigma_\alpha^2}(\boldsymbol{\theta}_0) = \sum_{i=1}^{g} \frac{{m_i}^2}{2(\dot{\sigma_e}^2 + m_i\dot{\sigma_\alpha}^2)^2}(\alpha_i^2 - \dot{\sigma_\alpha}^2),$$

$$\xi_{\beta_2}(\boldsymbol{\theta}_0) = \frac{1}{\dot{\sigma_e}^2}\sum_{i=1}^{g}\sum_{j=1}^{m_i}(x_{ij} - \bar{x}_{i.})e_{ij} \qquad \text{and}$$

$$\xi_{\sigma_e^2}(\boldsymbol{\theta}_0) = \frac{1}{2\dot{\sigma_e}^4}\sum_{i=1}^{g}\sum_{j=1}^{m_i}(e_{ij}^2 - \dot{\sigma_e}^2).$$

Next, we can obtain the following results

$$\lim_{g,m_L\to\infty} \mathrm{E}\left[g^{-\frac{1}{2}}\sum_{i=1}^{g}\left\{\frac{m_i}{(\dot{\sigma_e}^2 + m_i\dot{\sigma_\alpha}^2)} - \frac{1}{\dot{\sigma_\alpha}^2}\right\}\alpha_i\right] = 0,$$

$$\lim_{g,m_L\to\infty} \mathrm{Var}\left[g^{-\frac{1}{2}}\sum_{i=1}^{g}\left\{\frac{m_i}{(\dot{\sigma_e}^2 + m_i\dot{\sigma_\alpha}^2)} - \frac{1}{\dot{\sigma_\alpha}^2}\right\}\alpha_i\right] = \lim_{g,m_L\to\infty} g^{-1}\sum_{i=1}^{g}\frac{\dot{\sigma_e}^4}{\dot{\sigma_\alpha}^2(\dot{\sigma_e}^2 + m_i\dot{\sigma_\alpha}^2)^2} = o(1),$$

$$\lim_{g,m_L\to\infty} \mathrm{E}\left[g^{-\frac{1}{2}}\sum_{i=1}^{g}\left\{\frac{m_i^2}{2(\dot{\sigma_e}^2 + m_i\dot{\sigma_\alpha}^2)^2} - \frac{1}{2\dot{\sigma_\alpha}^4}\right\}(\alpha_i^2 - \dot{\sigma_\alpha}^2)\right] = 0 \qquad \text{and}$$

$$\lim_{g,m_L\to\infty} \mathrm{Var}\left[g^{-\frac{1}{2}}\sum_{i=1}^{g}\left\{\frac{m_i^2}{2(\dot{\sigma_e}^2 + m_i\dot{\sigma_\alpha}^2)^2} - \frac{1}{2\dot{\sigma_\alpha}^4}\right\}(\alpha_i^2 - \dot{\sigma_\alpha}^2)\right]$$
$$= \lim_{g,m_L\to\infty} g^{-1}\sum_{i=1}^{g}\frac{(\dot{\sigma_e}^4 + 2m_i\dot{\sigma_\alpha}^2\dot{\sigma_e}^2)^2}{4\dot{\sigma_\alpha}^8(\dot{\sigma_e}^2 + m_i\dot{\sigma_\alpha}^2)^4}\left\{\mathrm{E}(\alpha_i^4) - \dot{\sigma_\alpha}^4\right\}$$
$$= o(1)$$

These results allow us to rewrite the components of $\xi(\boldsymbol{\theta}_0)$ as

$$\begin{aligned}
\xi_{\beta_0}(\boldsymbol{\theta}_0) &= \frac{1}{\dot{\sigma_\alpha}^2}\sum_{i=1}^{g}\alpha_i,\\
\xi_{\beta_1}(\boldsymbol{\theta}_0) &= \frac{1}{\dot{\sigma_\alpha}^2}\sum_{i=1}^{g}\bar{x}_{i.}\alpha_i,\\
\xi_{\sigma_\alpha^2}(\boldsymbol{\theta}_0) &= \frac{1}{2\dot{\sigma_\alpha}^4}\sum_{i=1}^{g}(\alpha_i^2 - \dot{\sigma_\alpha}^2),\\
\xi_{\beta_2}(\boldsymbol{\theta}_0) &= \frac{1}{\dot{\sigma_e}^2}\sum_{i=1}^{g}\sum_{j=1}^{m_i}(x_{ij} - \bar{x}_{i.})e_{ij} \qquad \text{and}\\
\xi_{\sigma_e^2}(\boldsymbol{\theta}_0) &= \frac{1}{2\dot{\sigma_e}^4}\sum_{i=1}^{g}\sum_{j=1}^{m_i}(e_{ij}^2 - \dot{\sigma_e}^2).
\end{aligned} \tag{3.2.10}$$

Let

$$\xi(\boldsymbol{\theta}_0) = \begin{bmatrix} \xi^{(b)}(\boldsymbol{\theta}_0) \\ \xi^{(w)}(\boldsymbol{\theta}_0) \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{g} \xi_i^{(b)}(\boldsymbol{\theta}_0) \\ \sum_{i=1}^{g}\sum_{j=1}^{m_i} \xi_{ij}^{(w)}(\boldsymbol{\theta}_0) \end{bmatrix}. \tag{3.2.11}$$

From (3.2.10), we find that $\xi^{(b)}(\boldsymbol{\theta}_0)$ only depends on the random effect, while $\xi^{(w)}(\boldsymbol{\theta}_0)$ only depends on the error effect. Also, It is clear that $\xi^{(b)}(\boldsymbol{\theta}_0)$ and $\xi^{(w)}(\boldsymbol{\theta}_0)$ are independent. In what follows, We use these results to compute the variance of $\xi(\boldsymbol{\theta}_0)$.

3.2.2 Computing G_n, the variance of the approximate ML estimating equation at $\boldsymbol{\omega}_0$

Using (3.2.10), we calculate the variance of $\mathbf{K}^{-\frac{1}{2}}\xi(\boldsymbol{\theta}_0)$ as

$$\mathrm{Var}\left\{g^{-\frac{1}{2}}\xi_{\beta_0}(\boldsymbol{\theta}_0)\right\} = \frac{1}{\dot{\sigma_\alpha}^2}, \qquad \mathrm{Cov}\left\{g^{-\frac{1}{2}}\xi_{\beta_0}(\boldsymbol{\theta}_0), g^{-\frac{1}{2}}\xi_{\beta_1}(\boldsymbol{\theta}_0)\right\} = \frac{1}{g}\sum_{i=1}^{g}\frac{\bar{x}_{i.}}{\dot{\sigma_\alpha}^2},$$

$$\mathrm{Cov}\left\{g^{-\frac{1}{2}}\xi_{\beta_0}(\boldsymbol{\theta}_0), g^{-\frac{1}{2}}\xi_{\dot{\sigma_\alpha}^2}(\boldsymbol{\theta}_0)\right\} = \frac{\mathrm{E}(\alpha_i^3)}{2\dot{\sigma_\alpha}^6}, \qquad \mathrm{Var}\left\{g^{-\frac{1}{2}}\xi_{\beta_1}(\boldsymbol{\theta}_0)\right\} = \frac{1}{g}\sum_{i=1}^{g}\frac{\bar{x}_{i.}^2}{\dot{\sigma_\alpha}^2},$$

$$\mathrm{Cov}\left\{g^{-\frac{1}{2}}\xi_{\beta_1}(\boldsymbol{\theta}_0), g^{-\frac{1}{2}}\xi_{\dot{\sigma_\alpha}^2}(\boldsymbol{\omega}_0)\right\} = \frac{1}{g}\sum_{i=1}^{g}\frac{\bar{x}_{i.}\,\mathrm{E}(\alpha_i^3)}{2\dot{\sigma_\alpha}^6}, \qquad \mathrm{Var}\left\{g^{-\frac{1}{2}}\xi_{\sigma_\alpha^2}(\boldsymbol{\theta}_0)\right\} = \frac{\mathrm{E}\,\alpha_i^4 - \dot{\sigma_\alpha}^4}{4\dot{\sigma_\alpha}^8},$$

$$\mathrm{Var}\left\{n^{-\frac{1}{2}}\xi_{\beta_2}(\boldsymbol{\theta}_0)\right\} = \frac{S_w^x}{n\dot{\sigma_e}^2}, \qquad \mathrm{Cov}\left\{n^{-\frac{1}{2}}\xi_{\beta_2}(\boldsymbol{\theta}_0), n^{-\frac{1}{2}}\xi_{\sigma_e^2}(\boldsymbol{\omega}_0)\right\} = 0$$

and $$\mathrm{Var}\left\{n^{-\frac{1}{2}}\xi_{\sigma_e^2}(\boldsymbol{\theta}_0)\right\} = \frac{\mathrm{E}\,e_{ij}^4 - \sigma_e^4}{4\dot{\sigma_e}^8}.$$

Also, from (3.2.10), we see that $\xi^{(b)}(\boldsymbol{\theta}_0)$ only depends on α_i and $\xi^{(w)}(\boldsymbol{\theta}_0)$ only depends on e_{ij}. Since α_i and e_{ij} are independent, then we have

$$\mathbf{G}_n = \mathrm{Var}\begin{bmatrix} g^{-\frac{1}{2}}\xi^{(b)}(\boldsymbol{\theta}_0) \\ n^{-\frac{1}{2}}\xi^{(w)}(\boldsymbol{\theta}_0) \end{bmatrix} = \begin{bmatrix} \mathrm{Var}\left\{g^{-\frac{1}{2}}\xi^{(b)}(\boldsymbol{\theta}_0)\right\} & \mathbf{0} \\ \mathbf{0} & \mathrm{Var}\left\{n^{-\frac{1}{2}}\xi^{(w)}(\boldsymbol{\theta}_0)\right\} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{\dot{\sigma_\alpha}^2} & \frac{1}{g}\sum_{i=1}^{g}\frac{\bar{x}_{i.}}{\dot{\sigma_\alpha}^2} & \frac{\mathrm{E}(\alpha_i^3)}{2\dot{\sigma_\alpha}^6} & 0 & 0 \\ \frac{1}{g}\sum_{i=1}^{g}\frac{\bar{x}_{i.}}{\dot{\sigma_\alpha}^2} & \frac{1}{g}\sum_{i=1}^{g}\frac{\bar{x}_{i.}^2}{\dot{\sigma_\alpha}^2} & \frac{1}{g}\sum_{i=1}^{g}\frac{\bar{x}_{i.}\,\mathrm{E}(\alpha_i^3)}{2\dot{\sigma_\alpha}^6} & 0 & 0 \\ \frac{\mathrm{E}(\alpha_i^3)}{2\dot{\sigma_\alpha}^6} & \frac{1}{g}\sum_{i=1}^{g}\frac{\bar{x}_{i.}\,\mathrm{E}(\alpha_i^3)}{2\dot{\sigma_\alpha}^6} & \frac{\mathrm{E}(\alpha_i^4)-\dot{\sigma_\alpha}^4}{4\dot{\sigma_\alpha}^8} & 0 & 0 \\ 0 & 0 & 0 & \frac{S_w^x}{n\dot{\sigma_e}^2} & 0 \\ 0 & 0 & 0 & 0 & \frac{\mathrm{E}(e_{ij}^4)-\sigma_e^4}{4\dot{\sigma_e}^8} \end{bmatrix}.$$

From Condition B3 that $\lim_{g\to\infty}\frac{1}{g}\sum_{i=1}^{g}\bar{x}_{i.} = \mu_x$, $\lim_{g\to\infty}\frac{1}{g}\sum_{i=1}^{g}\bar{x}_{i.}^2 = \mu_x^2+\tau_x^2$ and $\lim_{g,m_L\to\infty}\frac{1}{n}S_w^x = \sigma_x^2$. Thus, we obtain

$$
\mathbf{G} = \lim_{g\to\infty}\lim_{m_L\to\infty}\mathbf{G}_n = \begin{bmatrix} \frac{1}{\dot{\sigma_\alpha}^2} & \frac{\mu_x}{\dot{\sigma_\alpha}^2} & \frac{\mathrm{E}(\alpha_i^3)}{2\dot{\sigma_\alpha}^6} & 0 & 0 \\ \frac{\mu_x}{\dot{\sigma_\alpha}^2} & \frac{\mu_x^2+\tau_x^2}{\dot{\sigma_\alpha}^2} & \frac{\mu_x\,\mathrm{E}(\alpha_i^3)}{2\dot{\sigma_\alpha}^6} & 0 & 0 \\ \frac{\mathrm{E}(\alpha_i^3)}{2\dot{\sigma_\alpha}^6} & \frac{\mu_x\,\mathrm{E}(\alpha_i^3)}{2\dot{\sigma_\alpha}^6} & \frac{\mathrm{E}(\alpha_i^4)-\dot{\sigma_\alpha}^4}{4\dot{\sigma_\alpha}^8} & 0 & 0 \\ 0 & 0 & 0 & \frac{\sigma_x^2}{\dot{\sigma_e}^2} & 0 \\ 0 & 0 & 0 & 0 & \frac{\mathrm{E}(e_{ij}^4)-\dot{\sigma_e}^4}{4\dot{\sigma_e}^8} \end{bmatrix} = \begin{bmatrix} \mathbf{G}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_2 \end{bmatrix}, \tag{3.2.12}
$$

where all the elements of both $\mathbf{G}_1$ and $\mathbf{G}_2$ are finite. Hence, we have $\|\mathbf{G}_1\| < \infty$ and $\|\mathbf{G}_2\| < \infty$. In summary, we find that starting from the angle of partitioning the parameters into within cluster and between cluster parameters, we are able to obtain an elegant form for matrix $\mathbf{G}$ as a block diagonal matrix formed from $\mathbf{G}_1$ and $\mathbf{G}_2$. Otherwise, if we have started from the angle of the fixed parameters $\boldsymbol{\beta}$ and the random parameters $\boldsymbol{\theta}$ directly, then we will find that the approximations of the ML estimating equations with respect to $\boldsymbol{\beta}_0$ and $\boldsymbol{\theta}_0$ are not independent, and thus $\mathbf{G}$ possess no simple structure that facilitate further algebraic developments.

3.2.3 Computing the derivative matrix from the ML estimating equations

From (3.1.5), we write $\psi^{(b)}(\boldsymbol{\omega}) = \sum_{i=1}^{g}\psi_i^{(b)}(\boldsymbol{\omega})$ and $\psi^{(w)}(\boldsymbol{\omega}) = \sum_{i=1}^{g}\sum_{j=1}^{m_i}\psi_{ij}^{(w)}(\boldsymbol{\omega})$, where the components of $\psi_i^{(b)}(\boldsymbol{\omega})$ and $\psi_{ij}^{(w)}(\boldsymbol{\omega})$ are defined as

$$
\begin{aligned}
l_{i\beta_0}(\boldsymbol{\omega}) &= \frac{m_i}{(\sigma_e^2+m_i\sigma_\alpha^2)}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1) \\
l_{i\beta_1}(\boldsymbol{\omega}) &= \frac{m_i\bar{x}_{i.}}{(\sigma_e^2+m_i\sigma_\alpha^2)}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1) \\
l_{i\sigma_\alpha^2}(\boldsymbol{\omega}) &= -\frac{1}{2}\frac{m_i}{(\sigma_e^2+m_i\sigma_\alpha^2)} + \frac{{m_i}^2}{2(\sigma_e^2+m_i\sigma_\alpha^2)^2}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1)^2 \\
l_{ij\beta_2}(\boldsymbol{\omega}) &= \frac{1}{\sigma_e^2}S_{ijw}^{xy} - \frac{1}{\sigma_e^2}S_{ijw}^{x}\beta_2 \\
l_{ij\sigma_e^2}(\boldsymbol{\omega}) &= -\frac{1}{2m_i(\sigma_e^2+m_i\sigma_\alpha^2)} - \frac{n-g}{2n\sigma_e^2} + \frac{1}{2\sigma_e^4}(S_{ijw}^{y} - 2S_{ijw}^{xy}\beta_2 + S_{ijw}^{x}\beta_2^2) \\
&\quad + \frac{1}{2(\sigma_e^2+m_i\sigma_\alpha^2)^2}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1)^2.
\end{aligned}
$$

Next, let

$$
\nabla\psi(\boldsymbol{\omega}) = \begin{bmatrix} \nabla\psi^{(b)}(\boldsymbol{\omega}) \\ \nabla\psi^{(w)}(\boldsymbol{\omega}) \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{g}\nabla\psi_i^{(b)}(\boldsymbol{\omega}) \\ \sum_{i=1}^{g}\sum_{j=1}^{m_i}\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega}) \end{bmatrix},
$$

where

$$\nabla\psi_i^{(b)}(\boldsymbol{\omega}) = \begin{bmatrix} l_{i\beta_0\beta_0}(\boldsymbol{\omega}) & l_{i\beta_0\beta_1}(\boldsymbol{\omega}) & l_{i\beta_0\sigma_\alpha^2}(\boldsymbol{\omega}) & l_{i\beta_0\beta_2}(\boldsymbol{\omega}) & l_{i\beta_0\sigma_e^2}(\boldsymbol{\omega}) \\ l_{i\beta_1\beta_0}(\boldsymbol{\omega}) & l_{i\beta_1\beta_1}(\boldsymbol{\omega}) & l_{i\beta_1\sigma_\alpha^2}(\boldsymbol{\omega}) & l_{i\beta_1\beta_2}(\boldsymbol{\omega}) & l_{i\beta_1\sigma_e^2}(\boldsymbol{\omega}) \\ l_{i\sigma_e^2\beta_0}(\boldsymbol{\omega}) & l_{i\sigma_e^2\beta_1}(\boldsymbol{\omega}) & l_{i\sigma_e^2\sigma_\alpha^2}(\boldsymbol{\omega}) & l_{i\sigma_e^2\beta_2}(\boldsymbol{\omega}) & l_{i\sigma_e^2\sigma_e^2}(\boldsymbol{\omega}) \end{bmatrix},$$

and

$$\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega}) = \begin{bmatrix} l_{i\beta_2\beta_0}(\boldsymbol{\omega}) & l_{i\beta_2\beta_1}(\boldsymbol{\omega}) & l_{i\beta_2\sigma_\alpha^2}(\boldsymbol{\omega}) & l_{i\beta_2\beta_2}(\boldsymbol{\omega}) & l_{i\beta_2\sigma_e^2}(\boldsymbol{\omega}) \\ l_{i\sigma_e^2\beta_0}(\boldsymbol{\omega}) & l_{i\sigma_e^2\beta_1}(\boldsymbol{\omega}) & l_{i\sigma_e^2\sigma_\alpha^2}(\boldsymbol{\omega}) & l_{i\sigma_e^2\beta_2}(\boldsymbol{\omega}) & l_{i\sigma_e^2\sigma_e^2}(\boldsymbol{\omega}) \end{bmatrix}.$$

We peoceed to compute each of these terms. For $\frac{\partial l_{i\beta_0}(\boldsymbol{\omega})}{\partial\boldsymbol{\omega}}$, the first row in $\nabla\psi_i^{(b)}(\boldsymbol{\omega})$, we have

$$\begin{aligned} &l_{i\beta_0\beta_0}(\boldsymbol{\omega}) = \frac{-m_i}{(\sigma_e^2+m_i\sigma_\alpha^2)}, \qquad l_{i\beta_0\beta_1}(\boldsymbol{\omega}) = \frac{-m_i\bar{x}_{i.}}{(\sigma_e^2+m_i\sigma_\alpha^2)}, \qquad l_{i\beta_0\sigma_\alpha^2}(\boldsymbol{\omega}) = \frac{-{m_i}^2}{(\sigma_e^2+m_i\sigma_\alpha^2)^2}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1), \\ &l_{i\beta_0\beta_2}(\boldsymbol{\omega}) = 0, \qquad \text{and} \quad l_{i\beta_0\sigma_e^2}(\boldsymbol{\omega}) = \frac{-m_i}{(\sigma_e^2+m_i\sigma_\alpha^2)^2}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1). \end{aligned} \tag{3.2.13}$$

For $\frac{\partial l_{i\beta_1}(\boldsymbol{\omega})}{\partial\boldsymbol{\omega}}$, the second row in $\nabla\psi_i^{(b)}(\boldsymbol{\omega})$, we have

$$\begin{aligned} &l_{i\beta_1\beta_0}(\boldsymbol{\omega}) = \frac{-m_i\bar{x}_{i.}}{(\sigma_e^2+m_i\sigma_\alpha^2)}, \qquad l_{i\beta_1\beta_1}(\boldsymbol{\omega}) = \frac{-m_i\bar{x}_{i.}^2}{(\sigma_e^2+m_i\sigma_\alpha^2)}, \qquad l_{i\beta_1\sigma_\alpha^2}(\boldsymbol{\omega}) = \frac{-{m_i}^2\bar{x}_{i.}}{(\sigma_e^2+m_i\sigma_\alpha^2)^2}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1), \\ &l_{i\beta_1\beta_2}(\boldsymbol{\omega}) = 0, \qquad \text{and} \quad l_{i\beta_1\sigma_e^2}(\boldsymbol{\omega}) = \frac{-m_i\bar{x}_{i.}}{(\sigma_e^2+m_i\sigma_\alpha^2)^2}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1). \end{aligned} \tag{3.2.14}$$

For $\frac{\partial l_{i\sigma_\alpha^2}(\boldsymbol{\omega})}{\partial\boldsymbol{\omega}}$,the third row in $\nabla\psi_i^{(b)}(\boldsymbol{\omega})$,we have

$$\begin{aligned} &l_{i\sigma_\alpha^2\beta_0}(\boldsymbol{\omega}) = \frac{-{m_i}^2}{(\sigma_e^2+m_i\sigma_\alpha^2)^2}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1), \qquad l_{i\sigma_\alpha^2\beta_1}(\boldsymbol{\omega}) = \frac{-{m_i}^2\bar{x}_{i.}}{(\sigma_e^2+m_i\sigma_\alpha^2)^2}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1), \\ &l_{i\sigma_\alpha^2\sigma_\alpha^2}(\boldsymbol{\omega}) = \frac{{m_i}^2}{2(\sigma_e^2+m_i\sigma_\alpha^2)^2} - \frac{{m_i}^3}{(\sigma_e^2+m_i\sigma_\alpha^2)^3}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1)^2, \\ &l_{i\sigma_\alpha^2\beta_2}(\boldsymbol{\omega}) = 0, \qquad \text{and} \qquad l_{i\sigma_\alpha^2\sigma_e^2}(\boldsymbol{\omega}) = \frac{m_i}{2(\sigma_e^2+m_i\sigma_\alpha^2)^2} - \frac{{m_i}^2}{(\sigma_e^2+m_i\sigma_\alpha^2)^3}(\bar{y}_{i.}-\beta_0-\bar{x}_{i.}\beta_1)^2. \end{aligned} \tag{3.2.15}$$

For $\frac{\partial l_{ij\beta_2}(\boldsymbol{\omega})}{\partial\boldsymbol{\omega}}$, the first row in$\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega})$, we have

$$l_{ij\beta_2\beta_0}(\boldsymbol{\omega}) = l_{ij\beta_2\beta_1}(\boldsymbol{\omega}) = l_{ij\beta_2\sigma_\alpha^2}(\boldsymbol{\omega}) = 0, \quad l_{ij\beta_2\beta_2}(\boldsymbol{\omega}) = \frac{-1}{\sigma_e^2}S_{ijw}^x, \quad \text{and} \quad l_{ij\beta_2\sigma_e^2}(\boldsymbol{\omega}) = \frac{-1}{\sigma_e^4}(S_{ijw}^{xy} - S_{ijw}^x\beta_2). \tag{3.2.16}$$

For $\frac{\partial l_{ij\sigma_e^2}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}}$, the second row in$\nabla \psi_{ij}^{(w)}(\boldsymbol{\omega})$, we have

$$
\begin{aligned}
& l_{ij\sigma_e^2\beta_0}(\boldsymbol{\omega}) = \frac{-1}{(\sigma_e^2 + m_i\sigma_\alpha^2)^2}(\bar{y}_{i.} - \beta_0 - \bar{x}_{i.}\beta_1), \quad l_{ij\sigma_e^2\beta_1}(\boldsymbol{\omega}) = \frac{-\bar{x}_{i.}}{(\sigma_e^2 + m_i\sigma_\alpha^2)^2}(\bar{y}_{i.} - \beta_0 - \bar{x}_{i.}\beta_1), \\
& l_{ij\sigma_e^2\sigma_\alpha^2}(\boldsymbol{\omega}) = \frac{1}{2(\sigma_e^2 + m_i\sigma_\alpha^2)^2} - \frac{m_i}{(\sigma_e^2 + m_i\sigma_\alpha^2)^3}(\bar{y}_{i.} - \beta_0 - \bar{x}_{i.}\beta_1)^2, \quad l_{ij\sigma_e^2\beta_2}(\boldsymbol{\omega}) = \frac{-1}{\sigma_e^4}(S_{ijw}^{xy} + S_{ijw}^{x}\beta_2), \\
& \text{and} \quad l_{ij\sigma_e^2\sigma_e^2}(\boldsymbol{\omega}) = \frac{1}{2m_i(\sigma_e^2 + m_i\sigma_\alpha^2)^2} + \frac{n-g}{2n\sigma_e^4} - \frac{1}{\sigma_e^6}(S_{ijw}^{y} - 2S_{ijw}^{xy}\beta_2 + S_{ijw}^{x}\beta_2^2) \\
& \qquad\qquad - \frac{1}{(\sigma_e^2 + m_i\sigma_\alpha^2)^3}(\bar{y}_{i.} - \beta_0 - \bar{x}_{i.}\beta_1)^2.
\end{aligned}
\tag{3.2.17}
$$

We use the above terms to calculate the expected derivative matrix

$$
\mathrm{E}\left\{\nabla\psi(\boldsymbol{\omega})\right\} = \begin{bmatrix} \sum_{i=1}^{g} \mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega})\right\} \\ \sum_{i=1}^{g}\sum_{j=1}^{m_i} \mathrm{E}\left\{\psi_{ij}^{(w)}(\boldsymbol{\omega})\right\} \end{bmatrix}.
$$

For $\mathrm{E}\left\{\frac{\partial l_{i\beta_0}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}}\right\}$, the first row of $\mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega})\right\}$, we can show that

$$
\begin{aligned}
& \mathrm{E}\left\{l_{i\beta_0\beta_0}(\boldsymbol{\omega})\right\} = \frac{-m_i}{(\sigma_e^2 + m_i\sigma_\alpha^2)}, \quad \mathrm{E}\left\{l_{i\beta_0\beta_1}(\boldsymbol{\omega})\right\} = \frac{-m_i\bar{x}_{i.}}{(\sigma_e^2 + m_i\sigma_\alpha^2)}, \\
& \mathrm{E}\left\{l_{i\beta_0\sigma_\alpha^2}(\boldsymbol{\omega})\right\} = \frac{-{m_i}^2}{(\sigma_e^2 + m_i\sigma_\alpha^2)^2}\left\{(\dot{\beta}_0 - \beta_0) + \bar{x}_{i.}(\dot{\beta}_2 - \beta_2)\right\}, \quad \mathrm{E}\left\{l_{i\beta_0\beta_2}(\boldsymbol{\omega})\right\} = 0, \\
& \text{and} \quad \mathrm{E}\left\{l_{i\beta_0\sigma_e^2}(\boldsymbol{\omega})\right\} = \frac{-m_i}{(\sigma_e^2 + m_i\sigma_\alpha^2)^2}\left\{(\dot{\beta}_0 - \beta_0) + \bar{x}_{i.}(\dot{\beta}_2 - \beta_2)\right\}.
\end{aligned}
\tag{3.2.18}
$$

When $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, we obtain

$$
\begin{aligned}
& \mathrm{E}\left\{l_{i\beta_0\beta_0}(\boldsymbol{\omega}_0)\right\} = \frac{-m_i}{({\dot{\sigma}_e}^2 + m_i{\dot{\sigma}_\alpha}^2)}, \quad \mathrm{E}\left\{l_{i\beta_0\beta_1}(\boldsymbol{\omega}_0)\right\} = \frac{-m_i\bar{x}_{i.}}{({\dot{\sigma}_e}^2 + m_i{\dot{\sigma}_\alpha}^2)}, \quad \mathrm{E}\left\{l_{i\beta_0\sigma_\alpha^2}(\boldsymbol{\omega}_0)\right\} = 0, \\
& \mathrm{E}\left\{l_{i\beta_0\beta_2}(\boldsymbol{\omega}_0)\right\} = 0, \quad \text{and} \quad \mathrm{E}\left\{l_{i\beta_0\sigma_e^2}(\boldsymbol{\omega}_0)\right\} = 0.
\end{aligned}
\tag{3.2.19}
$$

For $\mathrm{E}\left\{\frac{\partial l_{i\beta_1}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}}\right\}$, the second row of $\mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega})\right\}$, we can show that

$$
\begin{aligned}
& \mathrm{E}\left\{l_{i\beta_1\beta_0}(\boldsymbol{\omega})\right\} = \frac{-m_i\bar{x}_{i.}}{(\sigma_e^2 + m_i\sigma_\alpha^2)}, \qquad \mathrm{E}\left\{l_{i\beta_1\beta_1}(\boldsymbol{\omega})\right\} = \frac{-m_i\bar{x}_{i.}^2}{(\sigma_e^2 + m_i\sigma_\alpha^2)}, \\
& \mathrm{E}\left\{l_{i\beta_1\sigma_\alpha^2}(\boldsymbol{\omega})\right\} = \frac{-{m_i}^2\bar{x}_{i.}}{(\sigma_e^2 + m_i\sigma_\alpha^2)^2}\left\{(\dot{\beta}_0 - \beta_0) + \bar{x}_{i.}(\dot{\beta}_2 - \beta_2)\right\}, \\
& \mathrm{E}\left\{l_{i\beta_1\beta_2}(\boldsymbol{\omega})\right\} = 0, \qquad \text{and} \qquad \mathrm{E}\left\{l_{i\beta_1\sigma_e^2}(\boldsymbol{\omega})\right\} = \frac{-m_i\bar{x}_{i.}}{(\sigma_e^2 + m_i\sigma_\alpha^2)^2}\left\{(\dot{\beta}_0 - \beta_0) + \bar{x}_{i.}(\dot{\beta}_2 - \beta_2)\right\}.
\end{aligned}
\tag{3.2.20}
$$

When $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, we obtain

$$\mathrm{E}\left\{l_{i\beta_1\beta_0}(\boldsymbol{\omega}_0)\right\} = \frac{-m_i\bar{x}_{i.}}{(\dot{\sigma_e}^2 + m_i\dot{\sigma_\alpha}^2)}, \qquad \mathrm{E}\left\{l_{i\beta_1\beta_1}(\boldsymbol{\omega}_0)\right\} = \frac{-m_i\bar{x}_{i.}^2}{(\dot{\sigma_e}^2 + m_i\dot{\sigma_\alpha}^2)}, \qquad \mathrm{E}\left\{l_{i\beta_1\sigma_\alpha^2}(\boldsymbol{\omega}_0)\right\} = 0,$$
$$\mathrm{E}\left\{l_{i\beta_1\beta_2}(\boldsymbol{\omega}_0)\right\} = 0, \qquad \text{and} \qquad \mathrm{E}\left\{l_{i\beta_1\sigma_e^2}(\boldsymbol{\omega}_0)\right\} = 0. \tag{3.2.21}$$

Next, note from (3.2.4) and (3.2.6), we have

$$\mathrm{E}\{(\bar{y}_{i.} - \beta_0 - \bar{x}_{i.}\beta_1)^2\} = \{(\dot{\beta}_0 - \beta_0) + \bar{x}_{i.}(\dot{\beta}_2 - \beta_2)\}^2 + \dot{\sigma_\alpha}^2 + {m_i}^{-1}\dot{\sigma_e}^2, \tag{3.2.22}$$

and

$$\mathrm{E}(S^x_{ijw} - 2S^{xy}_{ijw}\beta_2 + S^x_{ijw}\beta_2^2) = S^x_{ijw}(\dot{\beta}_1 - \beta_1)^2 + {m_i}^{-1}(m_i - 1)\dot{\sigma_e}^2. \tag{3.2.23}$$

For $\mathrm{E}\left\{\frac{\partial l_{i\sigma_\alpha^2}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}}\right\}$, the third row of $\mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega})\right\}$, using (3.2.22) and (3.2.23), we can therefore obtain

$$\mathrm{E}\left\{l_{i\sigma_\alpha^2\beta_0}(\boldsymbol{\omega})\right\} = \frac{-{m_i}^2}{(\sigma_e^2 + m_i\sigma_\alpha^2)^2}\left\{(\dot{\beta}_0 - \beta_0) + \bar{x}_{i.}(\dot{\beta}_2 - \beta_2)\right\},$$
$$\mathrm{E}\left\{l_{i\sigma_\alpha^2\beta_1}(\boldsymbol{\omega})\right\} = \frac{-{m_i}^2\bar{x}_{i.}}{(\sigma_e^2 + m_i\sigma_\alpha^2)^2}\left\{(\dot{\beta}_0 - \beta_0) + \bar{x}_{i.}(\dot{\beta}_2 - \beta_2)\right\},$$
$$\mathrm{E}\left\{l_{i\sigma_\alpha^2\sigma_\alpha^2}(\boldsymbol{\omega})\right\} = \frac{{m_i}^2}{2(\sigma_e^2 + m_i\sigma_\alpha^2)^2} - \frac{{m_i}^3}{(\sigma_e^2 + m_i\sigma_\alpha^2)^3}\left[\left\{(\dot{\beta}_0 - \beta_0) + \bar{x}_{i.}(\dot{\beta}_2 - \beta_2)\right\}^2 + \dot{\sigma_\alpha}^2 + {m_i}^{-1}\dot{\sigma_e}^2\right],$$
$$\mathrm{E}\left\{l_{i\sigma_\alpha^2\beta_2}(\boldsymbol{\omega})\right\} = 0, \qquad \text{and}$$
$$\mathrm{E}\left\{l_{i\sigma_\alpha^2\sigma_e^2}(\boldsymbol{\omega})\right\} = \frac{m_i}{2(\sigma_e^2 + m_i\sigma_\alpha^2)^2} - \frac{{m_i}^2}{(\sigma_e^2 + m_i\sigma_\alpha^2)^3}\left[\left\{(\dot{\beta}_0 - \beta_0) + \bar{x}_{i.}(\dot{\beta}_2 - \beta_2)\right\}^2 + \dot{\sigma_\alpha}^2 + {m_i}^{-1}\dot{\sigma_e}^2\right]. \tag{3.2.24}$$

Moreover, when $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, we have

$$\mathrm{E}\left\{l_{i\sigma_\alpha^2\beta_0}(\boldsymbol{\omega}_0)\right\} = 0, \quad \mathrm{E}\left\{l_{i\sigma_\alpha^2\beta_1}(\boldsymbol{\omega}_0)\right\} = 0, \quad \mathrm{E}\left\{l_{i\sigma_\alpha^2\sigma_\alpha^2}(\boldsymbol{\omega}_0)\right\} = \frac{-{m_i}^2}{2(\dot{\sigma_e}^2 + m_i\dot{\sigma_\alpha}^2)^2}, \quad \mathrm{E}\left\{l_{i\sigma_\alpha^2\beta_2}(\boldsymbol{\omega}_0)\right\} = 0, \quad \text{and}$$
$$\mathrm{E}\left\{l_{i\sigma_\alpha^2\sigma_e^2}(\boldsymbol{\omega}_0)\right\} = \frac{-m_i}{2(\dot{\sigma_e}^2 + m_i\dot{\sigma_\alpha}^2)^2}. \tag{3.2.25}$$

For $\mathrm{E}\left\{\frac{\partial l_{ij\beta_2}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}}\right\}$, the first row of $\mathrm{E}\left\{\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega})\right\}$, we have

$$\begin{aligned}&\mathrm{E}\left\{l_{ij\beta_2\beta_0}(\boldsymbol{\omega})\right\}=\mathrm{E}\left\{l_{ij\beta_2\beta_1}(\boldsymbol{\omega})\right\}=\mathrm{E}\left\{l_{ij\beta_2\sigma_\alpha^2}(\boldsymbol{\omega})\right\}=0,\quad \mathrm{E}\left\{l_{ij\beta_2\beta_2}(\boldsymbol{\omega})\right\}=\frac{-1}{\sigma_e^2}S_{ijw}^x,\\ &\text{and}\quad \mathrm{E}\left\{l_{ij\beta_2\sigma_e^2}(\boldsymbol{\omega})\right\}=\frac{-1}{\sigma_e^4}S_{ijw}^x(\dot{\beta}_1-\beta_1).\end{aligned} \tag{3.2.26}$$

When $\boldsymbol{\omega}=\boldsymbol{\omega}_0$, we have

$$\begin{aligned}&\mathrm{E}\left\{l_{ij\beta_2\beta_0}(\boldsymbol{\omega}_0)\right\}=\mathrm{E}\left\{l_{ij\beta_2\beta_1}(\boldsymbol{\omega}_0)\right\}=\mathrm{E}\left\{l_{ij\beta_2\sigma_\alpha^2}(\boldsymbol{\omega}_0)\right\}=0,\quad \mathrm{E}\left\{l_{ij\beta_2\beta_2}(\boldsymbol{\omega}_0)\right\}=\frac{-1}{\dot{\sigma_e}^2}S_{ijw}^x\\ &\text{and}\quad \mathrm{E}\left\{l_{ij\beta_2\sigma_e^2}(\boldsymbol{\omega}_0)\right\}=0.\end{aligned} \tag{3.2.27}$$

Finally, for $\mathrm{E}\left\{\frac{\partial l_{ij\sigma_e^2}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}}\right\}$, the second row of $\mathrm{E}\left\{\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega})\right\}$, we can once again apply (3.2.22) and (3.2.23) to obtain

$$\begin{aligned}\mathrm{E}\left\{l_{ij\sigma_e^2\beta_0}(\boldsymbol{\omega})\right\}&=\frac{-1}{(\sigma_e^2+m_i\sigma_\alpha^2)^2}\left\{(\dot{\beta}_0-\beta_0)+\bar{x}_{i.}(\dot{\beta}_2-\beta_2)\right\},\\ \mathrm{E}\left\{l_{ij\sigma_e^2\beta_1}(\boldsymbol{\omega})\right\}&=\frac{-\bar{x}_{i.}}{(\sigma_e^2+m_i\sigma_\alpha^2)^2}\left\{(\dot{\beta}_0-\beta_0)+\bar{x}_{i.}(\dot{\beta}_2-\beta_2)\right\},\\ \mathrm{E}\left\{l_{ij\sigma_e^2\sigma_\alpha^2}(\boldsymbol{\omega})\right\}&=\frac{1}{2(\sigma_e^2+m_i\sigma_\alpha^2)^2}-\frac{m_i}{(\sigma_e^2+m_i\sigma_\alpha^2)^3}[\{(\dot{\beta}_0-\beta_0)+\bar{x}_{i.}(\dot{\beta}_2-\beta_2)\}^2+\dot{\sigma_\alpha}^2+m^{-1}\dot{\sigma_e}^2],\\ \mathrm{E}\left\{l_{ij\sigma_e^2\beta_2}(\boldsymbol{\omega})\right\}&=\frac{-1}{\sigma_e^4}S_{ijw}^x(\dot{\beta}_1-\beta_1),\qquad \text{and}\\ \mathrm{E}\left\{l_{ij\sigma_e^2\sigma_e^2}(\boldsymbol{\omega})\right\}&=\frac{1}{2m_i(\sigma_e^2+m_i\sigma_\alpha^2)^2}+\frac{n-g}{2n\sigma_e^4}-\frac{1}{\sigma_e^6}[S_{ijw}^x(\dot{\beta}_1-\beta_1)^2+{m_i}^{-1}(m_i-1)\dot{\sigma_e}^2]\\ &\quad-\frac{1}{(\sigma_e^2+m_i\sigma_\alpha^2)^3}\left[\left\{(\dot{\beta}_0-\beta_0)+\bar{x}_{i.}(\dot{\beta}_2-\beta_2)\right\}^2+\dot{\sigma_\alpha}^2+{m_i}^{-1}\dot{\sigma_e}^2\right].\end{aligned} \tag{3.2.28}$$

Furthermore, $\boldsymbol{\omega}=\boldsymbol{\omega}_0$, we have

$$\begin{aligned}&\mathrm{E}\left\{l_{ij\sigma_e^2\beta_0}(\boldsymbol{\omega}_0)\right\}=0,\quad \mathrm{E}\left\{l_{ij\sigma_e^2\beta_1}(\boldsymbol{\omega}_0)\right\}=0,\quad \mathrm{E}\left\{l_{ij\sigma_e^2\sigma_\alpha^2}(\boldsymbol{\omega}_0)\right\}=-\frac{1}{2(\dot{\sigma_e}^2+m_i\dot{\sigma_\alpha}^2)^2},\\ &\mathrm{E}\left\{l_{ij\sigma_e^2\beta_2}(\boldsymbol{\omega}_0)\right\}=0,\quad \text{and}\quad \mathrm{E}\left\{l_{ij\sigma_e^2\sigma_e^2}(\boldsymbol{\omega}_0)\right\}=\frac{-1}{2m_i(\dot{\sigma_e}^2+m_i\dot{\sigma_\alpha}^2)^2}-\frac{n-g}{2n\dot{\sigma_e}^4}.\end{aligned} \tag{3.2.29}$$

Combining all the results of (3.2.19), (3.2.21), (3.2.25), (3.2.27) and (3.2.29), we obtain

$$
\begin{aligned}
\mathbf{B}_n &= -\mathbf{K}^{-\frac{1}{2}} \operatorname{E} \nabla\psi(\boldsymbol{\omega}_0)\mathbf{K}^{-\frac{1}{2}} \\
&= \begin{bmatrix}
\frac{1}{g}\sum_{i=1}^{g} \frac{m_i}{(\dot\sigma_e^2+m_i\dot\sigma_\alpha^2)} & \frac{1}{g}\sum_{i=1}^{g} \frac{m_i\bar{x}_{i.}}{(\dot\sigma_e^2+m_i\dot\sigma_\alpha^2)} & 0 & 0 & 0 \\
\frac{1}{g}\sum_{i=1}^{g} \frac{m_i\bar{x}_{i.}}{(\dot\sigma_e^2+m_i\dot\sigma_\alpha^2)} & \frac{1}{g}\sum_{i=1}^{g} \frac{m_i\bar{x}_{i.}^2}{(\dot\sigma_e^2+m_i\dot\sigma_\alpha^2)} & 0 & 0 & 0 \\
0 & 0 & \frac{1}{g}\sum_{i=1}^{g} \frac{m_i^2}{2(\dot\sigma_e^2+m_i\dot\sigma_\alpha^2)^2} & 0 & \frac{1}{g}\sum_{i=1}^{g} \frac{m_i}{2(\dot\sigma_e^2+m_i\dot\sigma_\alpha^2)^2} \\
0 & 0 & 0 & \frac{S_w^x}{n\dot\sigma_e^2} & 0 \\
0 & 0 & \frac{1}{n}\sum_{i=1}^{g} \frac{m_i}{2(\dot\sigma_e^2+m_i\dot\sigma_\alpha^2)^2} & 0 & \frac{1}{n}\sum_{i=1}^{g} \frac{m_i}{2(\dot\sigma_e^2+m_i\dot\sigma_\alpha^2)^2} + \frac{(n-g)}{2n\dot\sigma_e^4}
\end{bmatrix}.
\end{aligned}
$$

Let $f(m_i) = \frac{m_i}{(\dot\sigma_e^2+m\dot\sigma_\alpha^2)}$ and $f = \frac{1}{\dot\sigma_\alpha^2}$, then we can show that

$$
\begin{aligned}
&\lim_{g,m_L\to\infty} \left|\frac{1}{g}\sum_{i=1}^{g} f(m_i)\bar{x}_{i.} - f\mu_x\right| = o(1) \qquad \text{and} \\
&\lim_{g,m_L\to\infty} \left|\frac{1}{g}\sum_{i=1}^{g} f(m_i)\bar{x}_{i.}^2 - f(\mu_x^2+\tau_x^2)\right| = o(1),
\end{aligned}
$$

where we use the fact that $\lim_{m_L\to\infty} |f(m_i) - f| \to 0$, $\lim_{g\to\infty}\sum_{i=1}^{g} |\bar{x}_{i.}| < \infty$, $\lim_{g\to\infty}\sum_{i=1}^{g} |\bar{x}_{i.}^2| < \infty$, $\lim_{g\to\infty}\sum_{i=1}^{g} \bar{x}_{i.} \to \mu_x$, and $\lim_{g\to\infty}\sum_{i=1}^{g} \bar{x}_{i.}^2 \to \mu_x^2 + \tau_x^2$ from Condition B3. It follows that

$$
\mathbf{B} = \lim_{g,m_L\to\infty} \mathbf{B}_n = \begin{bmatrix}
\frac{1}{\dot\sigma_\alpha^2} & \frac{\mu_x}{\dot\sigma_\alpha^2} & 0 & 0 & 0 \\
\frac{\mu_x}{\dot\sigma_\alpha^2} & \frac{\mu_x^2+\tau_x^2}{\dot\sigma_\alpha^2} & 0 & 0 & 0 \\
0 & 0 & \frac{1}{2\dot\sigma_\alpha^4} & 0 & 0 \\
0 & 0 & 0 & \frac{\sigma_x^2}{\dot\sigma_e^2} & 0 \\
0 & 0 & 0 & 0 & \frac{1}{2\dot\sigma_e^4}
\end{bmatrix} = \begin{bmatrix} \mathbf{B}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_2 \end{bmatrix}. \qquad (3.2.30)
$$

Here $\mathbf{B}_1$ is a 3×3 matrix and $\mathbf{B}_2$ is a 2×2 matrix. From (3.2.12) and (3.2.30), we can see that under normality , we obtain $\mathbf{G} = \mathbf{B}$.

3.3 Main results

The first step in establishing a central limit theorem for the maximum likelihood estimator $\hat{\boldsymbol{\omega}}$ is to find an appropriate normalizing matrix $\mathbf{K}$ that ensures that $\mathbf{K}^{-\frac{1}{2}}\{\psi(\boldsymbol{\omega}_0) - \xi(\boldsymbol{\omega}_0)\} = o_p(1)$ and that $\mathbf{K}^{-\frac{1}{2}}\xi(\boldsymbol{\omega}_0) \xrightarrow{D} N(\mathbf{0},\mathbf{G})$, where $\mathbf{G}$ is a positive definite covariance matrix. That is, $\mathbf{K}^{-\frac{1}{2}}\psi(\boldsymbol{\omega}_0)$ can be approximated by $\mathbf{K}^{-\frac{1}{2}}\xi(\boldsymbol{\omega}_0)$ which converges in distribution to a $N(\mathbf{0},\mathbf{G})$ distribution. Under condition A, we show that the appropriate normalizing matrix is the diagonal matrix $\mathbf{K} = \operatorname{diag}[g, g, g, n, n]$. This justifies focusing our attention on the neighborhood of $\boldsymbol{\omega}_0$ defined by $\mathcal{B} = \{\boldsymbol{\omega} : |\mathbf{K}^{\frac{1}{2}}(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega}_0)| \le M\}$, where M is a finite constant. The second step is to show that, uniformly on $\mathcal{B}$, the derivative matrix $\nabla\psi(\boldsymbol{\omega})$ of $\psi(\boldsymbol{\omega})$, pre- and post-multiplied by $\mathbf{K}^{-\frac{1}{2}}$, converges in probability to a positive definite matrix $-\mathbf{B}$. We can then establish that, uniformly on $\mathcal{B}$,

$$\mathbf{K}^{-\frac{1}{2}}\psi(\boldsymbol{\omega}) = \mathbf{K}^{-\frac{1}{2}}\xi(\boldsymbol{\omega}) - \mathbf{B}\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega} - \boldsymbol{\omega}_0) + o_p(1).$$

This uniform asymptotic linearity result allows us to deduce the existence of a consistent root to the estimating equation that satisfies the central limit theorem with asymptotic variance.

Theorem 3. *Suppose Condition B holds. Then there exists a solution $\hat{\boldsymbol{\omega}}$ to the estimating equations $\psi(\boldsymbol{\omega}) = 0$, satisfying $|\mathbf{K}^{\frac{1}{2}}(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega}_0)| = O_p(1)$ as $g \to \infty$ and $m_L \to \infty$, where $\mathbf{K} = \mathrm{diag}[g\mathbf{I}_{(b)}, n\mathbf{I}_{(w)}]$. Moreover,*

$$\mathbf{K}^{\frac{1}{2}}(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega}_0) = \mathbf{B}^{-1}\mathbf{K}^{-\frac{1}{2}}\xi(\boldsymbol{\omega}_0) + o_p(1),$$

whence

$$\mathbf{K}^{\frac{1}{2}}(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega}_0) \xrightarrow{D}, N(\mathbf{0}, \mathbf{B}^{-1}\mathbf{G}\mathbf{B}^{-1}),$$

where $\mathbf{B} = \lim_{g,m_L\to\infty} \mathbf{B}_n = -\mathbf{K}^{-\frac{1}{2}}\,\mathrm{E}\left\{\nabla\psi(\boldsymbol{\omega}_0)\right\}\mathbf{K}^{-\frac{1}{2}}$ *and* $\mathbf{G} = \lim_{g,m_L\to\infty} \mathbf{G}_n = \mathrm{Var}\left\{\mathbf{K}^{-\frac{1}{2}}\xi(\boldsymbol{\omega}_0)\right\}$.

To simplify the main proof, we establish a preliminary lemma and then break the main proof into a sequence of four additional lemmas. We establish a central limit theorem for the maximum likelihood estimator $\hat{\boldsymbol{\omega}}$ is to find an appropriate normalizing matrix $\mathbf{K}$ that ensures that $\mathbf{K}^{-\frac{1}{2}}\xi(\boldsymbol{\omega}_0) \xrightarrow{D} N(\mathbf{0}, \mathbf{G})$, where $\mathbf{G}$ is a positive definite covariance matrix.

Lemma 12. *Suppose Condition B holds. Then* $\lim_{g,m_L\to\infty} \mathbf{G}_n^{-\frac{1}{2}}\mathbf{K}^{-\frac{1}{2}}\xi(\boldsymbol{\theta}_0) \xrightarrow{D} N(\mathbf{0}, \mathbf{I})$.

Proof. The first step is to prove $\lim_{g\to\infty} g^{-\frac{1}{2}}\mathbf{G}_1^{-\frac{1}{2}}\xi^{(b)}(\boldsymbol{\theta}_0) \xrightarrow{D} N(\mathbf{0}, \mathbf{I}_{(b)})$. Suppose $\mathbf{a}$ is a fixed 3-vector satisfying $\mathbf{a}^T\mathbf{a} = 1$. From (3.2.11), we have $\mathbf{a}^T g^{-\frac{1}{2}}\mathbf{G}_1^{-\frac{1}{2}}\xi_1^{(b)}(\boldsymbol{\theta}_0), \ldots, \mathbf{a}^T g^{-\frac{1}{2}}\mathbf{G}_1^{-\frac{1}{2}}\xi_g^{(b)}(\boldsymbol{\theta}_0)$ are independent scalar random variables. As $g, m_L \to \infty$, $\mathbf{a}^T g^{-\frac{1}{2}}\mathbf{G}_1^{-\frac{1}{2}}\,\mathrm{E}\left\{\xi_i^{(b)}(\boldsymbol{\theta}_0)\right\} = 0$ and $\mathrm{Var}\left\{\mathbf{a}^T g^{-\frac{1}{2}}\mathbf{G}_1^{-\frac{1}{2}}\xi^{(b)}(\boldsymbol{\theta}_0)\right\} = \sum_{i=1}^{g}\mathrm{Var}\left\{\mathbf{a}^T g^{-\frac{1}{2}}\mathbf{G}_1^{-\frac{1}{2}}\xi_i^{(b)}(\boldsymbol{\theta}_0)\right\} = 1$. Also there exists a $\delta > 0$ such that

$$\begin{aligned}
&\lim_{g\to\infty}\frac{1}{1^{2+\delta}}\sum_{i=1}^{g}\mathrm{E}\,|\mathbf{a}^T g^{-\frac{1}{2}}\mathbf{G}_1^{-\frac{1}{2}}\xi_i^{(b)}(\boldsymbol{\theta}_0)|^{2+\delta} \\
&\le \lim_{g\to\infty} g^{-1-\frac{\delta}{2}}\|\mathbf{G}_1^{-\frac{1}{2}}\|^{2+\delta}\sum_{i=1}^{g}\mathrm{E}\,|\xi_i^{(b)}(\boldsymbol{\theta}_0)|^{2(1+\frac{\delta}{2})} \\
&= \lim_{g\to\infty} g^{-1-\frac{\delta}{2}}\|\mathbf{G}_1^{-\frac{1}{2}}\|^{2+\delta}\sum_{i=1}^{g}\mathrm{E}[\{\xi_{i\beta_0}^2(\boldsymbol{\theta}_0) + \xi_{i\beta_1}^2(\boldsymbol{\theta}_0) + \xi_{i\sigma_\alpha^2}^2(\boldsymbol{\theta}_0)\}^{1+\frac{\delta}{2}}] \\
&\le \lim_{g\to\infty} g^{-1-\frac{\delta}{2}}\|\mathbf{G}_1^{-\frac{1}{2}}\|^{2+\delta}\sum_{i=1}^{g}[\mathrm{E}\{\xi_{i\beta_0}^{2+\delta}(\boldsymbol{\theta}_0)\} + \mathrm{E}\{\xi_{i\beta_1}^{2+\delta}(\boldsymbol{\theta}_0)\} + \mathrm{E}\{\xi_{i\sigma_\alpha^2}^{2+\delta}(\boldsymbol{\theta}_0)\}].
\end{aligned}$$

From (3.2.10), we can show that

$$\xi^{2+\delta}_{i\beta_0}(\boldsymbol{\theta}_0) = (\frac{1}{\dot{\sigma_\alpha}^2})^{2+\delta}\alpha_i^{2+\delta},$$
$$\xi^{2+\delta}_{i\beta_1}(\boldsymbol{\theta}_0) = (\frac{\bar{x}_{i.}}{\dot{\sigma_\alpha}^2})^{2+\delta}\alpha_i^{2+\delta}, \qquad \text{and}$$
$$\xi^{2+\delta}_{i\sigma^2_\alpha}(\boldsymbol{\theta}_0) = (\frac{1}{2\dot{\sigma_\alpha}^4})^{2+\delta}(\alpha_i^2 - \dot{\sigma_\alpha}^2)^{2+\delta}.$$

Since $\mathrm{E}(\alpha_i^{4+\lambda}) \leq K < \infty$, $\frac{m_i}{(\dot{\sigma_e}^2 + m_i\dot{\sigma_\alpha}^2)} < \infty$ and using the fact that Condition B4 implies $\lim_{g\to\infty}\frac{1}{g}\sum_{i=1}^{g}\bar{x}_{i.}^{2+\delta} < \infty$, then we can show $\mathrm{E}\{\xi^{2+\delta}_{i\beta_0}(\boldsymbol{\theta}_0)\} + \mathrm{E}\{\xi^{2+\delta}_{i\beta_1}(\boldsymbol{\theta}_0)\} + \mathrm{E}\{\xi^{2+\delta}_{i\sigma^2_\alpha}(\boldsymbol{\theta}_0)\} < \infty$. Thus, we get

$$\lim_{g\to\infty} g^{-1-\frac{\delta}{2}}\|\mathbf{G}_1^{-\frac{1}{2}}\|^{2+\delta}\sum_{i=1}^{g}\left[\mathrm{E}\left\{\xi^{2+\delta}_{i\beta_0}(\boldsymbol{\theta}_0)\right\} + \mathrm{E}\left\{\xi^{2+\delta}_{i\beta_1}(\boldsymbol{\theta}_0)\right\} + \mathrm{E}\left\{\xi^{2+\delta}_{i\sigma^2_\alpha}(\boldsymbol{\theta}_0)\right\}\right] = O_p(g^{-\frac{\delta}{2}}) = o_p(1).$$

We find Lyapunov's condition holds. Consequently $\mathbf{a}^T g^{-\frac{1}{2}}\mathbf{G}_1^{-\frac{1}{2}}\xi^{(b)}(\boldsymbol{\theta}_0)$ converges in distribution to $N(0,1)$, as $g\to\infty$. The required result follows from the Cramer-Wold device.

The second step is to prove $\lim_{g,m_L\to\infty}\xi^{(w)}(\boldsymbol{\theta}_0) \xrightarrow{D} N(\mathbf{0}, \mathbf{I}_{(w)})$. Suppose $\mathbf{b}$ is a fixed 2-vector satisfying $\mathbf{b}^T\mathbf{b} = 1$. From (3.2.11), we have that $\mathbf{b}^T n^{-\frac{1}{2}}\mathbf{G}_2^{-\frac{1}{2}}\xi^{(w)}_{11}(\boldsymbol{\theta}_0), \ldots, \mathbf{b}^T n^{-\frac{1}{2}}\mathbf{G}_2^{-\frac{1}{2}}\xi^{(w)}_{gm_g}(\boldsymbol{\theta}_0)$ are independent scalar random variables. As $g\to\infty$ and $m_L\to\infty$, then $\mathbf{b}^T n^{-\frac{1}{2}}\mathbf{G}_2^{-\frac{1}{2}}\mathrm{E}\left\{\xi^{(w)}_{ij}(\boldsymbol{\theta}_0)\right\} = 0$ and $\mathrm{Var}\left\{\mathbf{b}^T n^{-\frac{1}{2}}\mathbf{G}_2^{-\frac{1}{2}}\xi^{(w)}(\boldsymbol{\theta}_0)\right\} = \sum_{i=1}^{g}\sum_{j=1}^{m_i}\mathrm{Var}\left\{\mathbf{b}^T n^{-\frac{1}{2}}\mathbf{G}_2^{-\frac{1}{2}}\xi^{(w)}_{ij}(\boldsymbol{\theta}_0)\right\} = 1$. Furthermore, there exists a $\delta > 0$ such that

$$\lim_{g,m_L\to\infty}\frac{1}{1^{2+\delta}}\sum_{i=1}^{g}\sum_{j=1}^{m_i}\mathrm{E}\,|\mathbf{b}^T n^{-\frac{1}{2}}\mathbf{G}_2^{-\frac{1}{2}}\xi^{(w)}_{ij}(\boldsymbol{\theta}_0)|^{2+\delta}$$
$$\leq \lim_{g,m_L\to\infty} n^{-1-\frac{\delta}{2}}\|\mathbf{G}_2^{-\frac{1}{2}}\|^{2+\delta}\sum_{i=1}^{g}\sum_{j=1}^{m_i}\mathrm{E}\,|\xi^{(w)}_{ij}(\boldsymbol{\theta}_0)|^{2(1+\frac{\delta}{2})}$$
$$\leq \lim_{g,m_L\to\infty} n^{-1-\frac{\delta}{2}}\|\mathbf{G}_2^{-\frac{1}{2}}\|^{2+\delta}\sum_{i=1}^{g}\sum_{j=1}^{m_i}[\mathrm{E}\{\xi^{2+\delta}_{ij\beta_2}(\boldsymbol{\theta}_0)\} + \mathrm{E}\{\xi^{2+\delta}_{ij\sigma^2_e}(\boldsymbol{\theta}_0)\}].$$

From (3.2.10) we can show that

$$\xi^{2+\delta}_{ij\beta_2}(\boldsymbol{\theta}_0) = \{\frac{1}{\dot{\sigma_e}^2}(x_{ij} - \bar{x}_{i.})\}^{2+\delta}e_{ij}^{2+\delta} \qquad \text{and} \qquad \xi^{2}_{ij\sigma^2_e}(\boldsymbol{\theta}_0) = (\frac{1}{2\dot{\sigma_e}^4})^{2+\delta}(e_{ij}^2 - \dot{\sigma_e}^2)^{2+\delta}.$$

Since $\mathrm{E}(e_{ij}^{4+\lambda}) < \infty$ and using the fact that Condition B2 implies $\lim_{g,m_L\to\infty}\frac{1}{n}\sum_{i=1}^{g}\sum_{j=1}^{m_i}(x_{ij} - \bar{x}_{i.})^{2+\delta} < \infty$, then we can show $\mathrm{E}\{\xi^{2+\delta}_{ij\beta_2}(\boldsymbol{\theta}_0)\} + \mathrm{E}\{\xi^{2+\delta}_{ij\sigma^2_e}(\boldsymbol{\theta}_0)\} < \infty$. Thus

$$\lim_{g,m_L\to\infty} n^{-1-\frac{\delta}{2}}\|\mathbf{G}_2^{-\frac{1}{2}}\|^{2+\delta}\sum_{i=1}^{g}\sum_{j=1}^{m_i}\left[\mathrm{E}\left\{\xi^{2+\delta}_{ij\beta_2}(\boldsymbol{\theta}_0)\right\} + \mathrm{E}\left\{\xi^{2+\delta}_{ij\sigma^2_e}(\boldsymbol{\theta}_0)\right\}\right] = O_p(n^{-\frac{\delta}{2}}) = o_p(1).$$

We find that Lyapunov's condition holds. Consequently $\mathbf{b}^T n^{-\frac{1}{2}}\mathbf{G}_2^{-\frac{1}{2}}\xi^{(w)}$ converges to $N(0,1)$, as

$g \to \infty$ and $m_L \to \infty$. The required result follows from the Cramer-Wold device. Since $\xi^{(b)}(\boldsymbol{\theta}_0)$ and $\xi^{(w)}(\boldsymbol{\theta}_0)$ are independent from (3.2.10), so we obtain $\mathbf{G}_n^{-\frac{1}{2}}\mathbf{K}^{-\frac{1}{2}}\xi(\boldsymbol{\theta}_0) \xrightarrow{D} N(\mathbf{0}, \mathbf{I})$. $\square$

We prove preliminary lemmas and then the asymptotic linearity result.

Lemma 13. *Suppose Condition B holds, then, for $i = 1, \ldots, g$ and $j = 1, \ldots, m_i$,*

$$\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\leq M} \mathrm{E}\,\|\nabla\psi_i^{(b)}(\boldsymbol{\omega})\|^2 < \infty \qquad \textit{and} \qquad \sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\leq M} \mathrm{E}\,\|\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega})\|^2 < \infty.$$

Proof. From (3.2.13)-(3.2.17), we can have that

$$\mathrm{E}(\bar{y}_{i.} - \beta_0 - \bar{x}_{i.}\beta_1)^2 = \left\{(\dot{\beta}_0 - \beta_0) + \bar{x}_{i.}(\dot{\beta}_1 - \beta_1)\right\}^2 + \dot{\sigma_\alpha}^2 + m_i^{-1}\dot{\sigma_e}^2 = O(1)$$

Therefore, we obtain

$$\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\leq M} \mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega})\nabla\psi_i^{(b)}(\boldsymbol{\omega})^T\right\} = \begin{bmatrix} O(1) & O(1) & O(1) & 0 & O(m_i^{-1}) \\ O(1) & O(1) & O(1) & 0 & O(m_i^{-1}) \\ O(1) & O(1) & O(1) & 0 & O(m_i^{-1}) \end{bmatrix}, \tag{3.3.1}$$

and

$$\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\leq M} \mathrm{E}\left\{\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega})\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega})^T\right\} = \begin{bmatrix} 0 & 0 & 0 & O(1) & O(1) \\ O(m_i^{-2}) & O(m_i^{-2}) & O(m_i^{-2}) & O(1) & O(1) \end{bmatrix}. \tag{3.3.2}$$

Thus, $\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\leq M} \mathrm{E}\,\|\nabla\psi_i^{(b)}(\boldsymbol{\omega})\|^2 < \infty$ and $\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\leq M} \mathrm{E}\,\|\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega})\|^2 < \infty$ as required. $\square$

Lemma 14. *Suppose Condition B holds. Let $t_k = (t_{k1}, t_{k2}, t_{k3}, t_{k4}, t_{k5})^T$. Decompose the cube $\mathcal{B} = \{\boldsymbol{\omega} : |\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega} - \boldsymbol{\omega}_0)| \leq M\}$ into smaller cubes $\mathcal{C} = \{\mathcal{C}(t_k)\}$, where M is a constant and $\boldsymbol{\omega} \in \mathcal{C}(t_k) = \{\boldsymbol{\omega} : |\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega} - \boldsymbol{\omega}_0)| \leq Mg^{-\frac{1}{4}}\}$. As $g \to \infty$ and $m_L \to \infty$, then*

$$\max_{1\leq k\leq N}\sup_{\boldsymbol{\omega}\in\mathcal{C}(t_k)} g^{-\frac{1}{2}}\sum_{i=1}^{g} \left|\left[\nabla\psi_i^{(b)}(\tilde{\boldsymbol{\omega}}) - \mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega}^*)\right\}\right](\boldsymbol{\omega} - t_k)\right| = o_p(1),$$
$$\max_{1\leq k\leq N}\sup_{\boldsymbol{\omega}\in\mathcal{C}(t_k)} n^{-\frac{1}{2}}\sum_{i=1}^{g}\sum_{j=1}^{m_i} \left|\left[\nabla\psi_{ij}^{(w)}(\tilde{\boldsymbol{\omega}}) - \mathrm{E}\left\{\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega}^*)\right\}\right](\boldsymbol{\omega} - t_k)\right| = o_p(1),$$

where $N = g^{\frac{1}{4}}$ cubes and $\tilde{\boldsymbol{\omega}}, \boldsymbol{\omega}^* \in \mathcal{C}(t_k)$.

Proof. Let $\|\mathbf{K}^{-\frac{1}{2}}\| \leq g^{-\frac{1}{2}}L$ for some constant $L > 0$. For the first equation in lemma 14, we have

$$\begin{aligned}
&\max_{1\le k\le N}\sup_{\boldsymbol{\omega}\in\mathcal{C}(t_k)} g^{-\frac{1}{2}}\sum_{i=1}^{g}|\left[\nabla\psi_i^{(b)}(\tilde{\boldsymbol{\omega}})-\mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega}^*)\right\}\right](\boldsymbol{\omega}-t_k)|\\
&\le \max_{1\le k\le N}\sup_{\boldsymbol{\omega}\in\mathcal{C}(t_k)} g^{-\frac{1}{2}}\sum_{i=1}^{g}\|\nabla\psi_i^{(b)}(\tilde{\boldsymbol{\omega}})-\mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega}^*)\right\}\|\|\mathbf{K}^{-\frac{1}{2}}\||\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-t_k)|\\
&\le \max_{1\le k\le N}\sup_{\boldsymbol{\omega}\in\mathcal{C}(t_k)} g^{-\frac{1}{4}}M\, g^{-\frac{1}{2}}\sum_{i=1}^{g}\|\nabla\psi_i^{(b)}(\tilde{\boldsymbol{\omega}})-\mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega}^*)\right\}\|\|\mathbf{K}^{-\frac{1}{2}}\|\\
&\le \max_{1\le k\le N}\sup_{\boldsymbol{\omega}\in\mathcal{C}(t_k)} g^{-1-\frac{1}{4}}ML\,\sum_{i=1}^{g}\{\|\nabla\psi_i^{(b)}(\tilde{\boldsymbol{\omega}})\|+\mathrm{E}\,\|\nabla\psi_i^{(b)}(\boldsymbol{\omega}^*)\|\}\\
&=O_p(g^{-\frac{1}{4}})\\
&=o_p(1),
\end{aligned}$$

where we use the fact that $\|\nabla\psi_i^{(b)}(\tilde{\boldsymbol{\omega}})\|=O_p(1)$ and $\mathrm{E}\,\|\nabla\psi_i^{(b)}(\boldsymbol{\omega}^*)\|=O_p(1)$ from (3.3.1).

For the second equation in lemma 14, we require a more delicate argument. We get

$$\begin{aligned}
&\max_{1\le k\le N}\sup_{\boldsymbol{\omega}\in\mathcal{C}(t_k)} n^{-\frac{1}{2}}\sum_{i=1}^{g}\sum_{j=1}^{m_i}|\left[\nabla\psi_{ij}^{(w)}(\tilde{\boldsymbol{\omega}})-\mathrm{E}\left\{\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega}^*)\right\}\right](\boldsymbol{\omega}-t_k)|\\
&\le \max_{1\le k\le n}\sup_{\boldsymbol{\omega}\in\mathcal{C}(t_k)} n^{-\frac{1}{2}}\sum_{i=1}^{g}\sum_{j=1}^{m_i}\Bigg[|\left\{\frac{\partial l_{ij\beta_2}(\tilde{\boldsymbol{\omega}})}{\partial\boldsymbol{\omega}}-\mathrm{E}\,\frac{\partial l_{ij\beta_2}(\boldsymbol{\omega}^*)}{\partial\boldsymbol{\omega}}\right\}(\boldsymbol{\omega}-t_k)|\\
&\quad+|\left\{\frac{\partial l_{ij\sigma_e^2}(\tilde{\boldsymbol{\omega}})}{\partial\boldsymbol{\omega}}-\mathrm{E}\,\frac{\partial l_{ij\sigma_e^2}(\boldsymbol{\omega}^*)}{\partial\boldsymbol{\omega}}\right\}(\boldsymbol{\omega}-t_k)|\Bigg]\\
&\le \max_{1\le k\le n}\sup_{\boldsymbol{\omega}\in\mathcal{C}(t_k)} n^{-\frac{1}{2}}\sum_{i=1}^{g}\sum_{j=1}^{m_i}\Big[|\{l_{ij\beta_2\beta_2}(\tilde{\boldsymbol{\omega}})-\mathrm{E}\,l_{ij\beta_2\beta_2}(\boldsymbol{\omega}^*)\}(\beta_2-t_{k4})|\\
&\quad+|\left\{l_{ij\beta_2\sigma_e^2}(\tilde{\boldsymbol{\omega}})-\mathrm{E}\,l_{ij\beta_2\sigma_e^2}(\boldsymbol{\omega}^*)\right\}(\sigma_e^2-t_{k5})+|\left\{l_{ij\sigma_e^2\beta_0}(\tilde{\boldsymbol{\omega}})-\mathrm{E}\,l_{ij\sigma_e^2\beta_0}(\boldsymbol{\omega}^*)\right\}(\beta_0-t_{k1})||\\
&\quad+|\left\{l_{ij\sigma_e^2\beta_1}(\tilde{\boldsymbol{\omega}})-\mathrm{E}\,l_{ij\sigma_e^2\beta_1}(\boldsymbol{\omega}^*)\right\}(\beta_1-t_{k2})|+|\left\{l_{ij\sigma_e^2\sigma_\alpha^2}(\tilde{\boldsymbol{\omega}})-\mathrm{E}\,l_{ij\sigma_e^2\sigma_\alpha^2}(\boldsymbol{\omega}^*)\right\}(\sigma_\alpha^2-t_{k3})||\\
&\quad+|\left\{l_{ij\sigma_e^2\beta_2}(\tilde{\boldsymbol{\omega}})-\mathrm{E}\,l_{ij\sigma_e^2\beta_2}(\boldsymbol{\omega}^*)\right\}(\beta_2-t_{k4})|+|\left\{l_{ij\sigma_e^2\sigma_e^2}(\tilde{\boldsymbol{\omega}})-\mathrm{E}\,l_{ij\sigma_e^2\sigma_e^2}(\boldsymbol{\omega}^*)\right\}(\sigma_e^2-t_{k5})|\Big],
\end{aligned}$$

where we have applied the results from (3.2.16), (3.2.17), (3.2.26) and (3.2.28). Next, We have that

$$|l_{ij\sigma_e^2\beta_0}(\tilde{\boldsymbol{\omega}})-\mathrm{E}\,l_{ij\sigma_e^2\beta_0}(\boldsymbol{\omega}^*)|=|\frac{-1}{(\tilde{\sigma}_e^2+m_i\tilde{\sigma}_\alpha^2)^2}(\bar{y}_{i.}-\tilde{\beta}_0-\bar{x}_{i.}\tilde{\beta}_1)-0|=\frac{1}{m_i^2}\left|\frac{\bar{y}_{i.}-\tilde{\beta}_0-\bar{x}_{i.}\tilde{\beta}_1}{(\tilde{\sigma}_\alpha^2+\frac{\tilde{\sigma}_e^2}{m_i})^2}\right|.$$

From Condition A4, we know $\max_{1\le i\le g}|\bar{x}_{i.}|<\infty$. Thus we obtain $|l_{ij\sigma_e^2\beta_0}(\tilde{\boldsymbol{\omega}})-\mathrm{E}\,l_{ij\sigma_e^2\beta_0}(\boldsymbol{\omega}^*)|=O_p(m_i^{-2})$. Similarly, we can show that $|l_{ij\sigma_e^2\beta_1}(\tilde{\boldsymbol{\omega}})-\mathrm{E}\,l_{ij\sigma_e^2\beta_1}(\boldsymbol{\omega}^*)|=O_p(m_i^{-2})$ and $|l_{ij\sigma_e^2\sigma_\alpha^2}(\tilde{\boldsymbol{\omega}})-\mathrm{E}\,l_{ij\sigma_e^2\sigma_\alpha^2}(\boldsymbol{\omega}^*)|=O_p(m_i^{-2})$ as $m_L\to\infty$. Since $\boldsymbol{\omega}\in\mathcal{C}(t_k)$, then we have $|\beta_0-t_{k1}|\le g^{-\frac{3}{4}}M$, $|\beta_1-t_{k2}|\le g^{-\frac{3}{4}}M$, $|\sigma_\alpha^2-t_{k3}|\le g^{-\frac{3}{4}}M$, $|\beta_2-t_{k4}|\le g^{-\frac{1}{4}}n^{-\frac{1}{2}}M$ and $|\sigma_e^2-t_{k5}|\le g^{-\frac{1}{4}}n^{-\frac{1}{2}}M$.

Thus, we obtain

$$
\begin{aligned}
&\max_{1\le k\le N}\sup_{\boldsymbol{\omega}\in\mathcal{C}(t_k)} n^{-\frac{1}{2}}\sum_{i=1}^{g}\sum_{j=1}^{m_i}\left|\left[\nabla\psi_{ij}^{(w)}(\tilde{\boldsymbol{\omega}})-\mathrm{E}\left\{\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega}^*)\right\}\right](\boldsymbol{\omega}-t_k)\right|\\
&=\min_{1\le i\le g}\{O(g^{-\frac{1}{4}}),O(n^{\frac{1}{2}}g^{-\frac{3}{4}}m_i^{-2})\}\\
&=o_p(1).
\end{aligned}
$$

□

Lemma 15. *Suppose Condition B holds. As $g\to\infty$ and $m_L\to\infty$, then we have*

$$
\begin{aligned}
&\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} g^{-\frac{1}{2}}\sum_{i=1}^{g}\left|\left[\mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega})\right\}-\mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega}_0)\right\}\right](\boldsymbol{\omega}-\boldsymbol{\omega}_0)\right|=o(1),\\
&\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} n^{-\frac{1}{2}}\sum_{i=1}^{g}\sum_{j=1}^{m_i}\left|\left[\mathrm{E}\left\{\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega})\right\}-\mathrm{E}\,\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega}_0)\right](\boldsymbol{\omega}-\boldsymbol{\omega}_0)\right|=o(1).
\end{aligned}
$$

Proof. For the first equation of lemma 15, we have

$$
\begin{aligned}
&\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} g^{-\frac{1}{2}}\sum_{i=1}^{g}\left|\left[\mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega})\right\}-\mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega}_0)\right\}\right](\boldsymbol{\omega}-\boldsymbol{\omega}_0)\right|\\
&\le\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} g^{-\frac{1}{2}}\sum_{i=1}^{g}\left\|\mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega})\right\}-\mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega}_0)\right\}\right\|\|\mathbf{K}^{-\frac{1}{2}}\||\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\\
&\le\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} g^{-1}LM\sum_{i=1}^{g}\left\|\mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega})\right\}-\mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega}_0)\right\}\right\|,
\end{aligned}
$$

where $\|\mathbf{K}^{-\frac{1}{2}}\|\le g^{-\frac{1}{2}}L$ and$L$ and M are both positive constant. For convenient calculation, we denote $\zeta_i=\frac{m_i}{(\sigma_e^2+m_i\sigma_\alpha^2)}$, $\dot{\zeta}_i=\frac{m_i}{(\dot{\sigma_e}^2+m_i\dot{\sigma_\alpha}^2)}$ and $h_i(\beta_0,\beta_1)=(\dot{\beta}_0-\beta_0)+\bar{x}_{i.}(\dot{\beta}_1-\beta_1)$. Since $|\mathbf{K}^{-\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M$ and Condition B4 implies that $\bar{x}_{i.}$ and $\bar{x}_{i.}^2$ are bounded as $g\to\infty$ and $m_L\to\infty$, then we can show that $|\beta_0-\dot{\beta}_0|\le g^{-\frac{1}{2}}M$, $|\beta_1-\dot{\beta}_1|\le g^{-\frac{1}{2}}M$, $|\sigma_\alpha^2-\dot{\sigma_\alpha}^2|\le g^{-\frac{1}{2}}M$, $|\beta_2-\dot{\beta}_2|\le n^{-\frac{1}{2}}M$, and $|\sigma_e^2-\dot{\sigma_e}^2|\le n^{-\frac{1}{2}}M$. As a result, we obtain

$$
\begin{aligned}
&|h_i(\beta_0,\beta_1)|=|(\dot{\beta}_0-\beta_0)+\bar{x}_{i.}(\dot{\beta}_1-\beta_1)|\le g^{-\frac{1}{2}}M,\\
&|\zeta_i-\dot{\zeta}_i|=\left|\frac{m_i}{(\sigma_e^2+m_i\sigma_\alpha^2)}-\frac{m_i}{(\dot{\sigma_e}^2+m_i\dot{\sigma_\alpha}^2)}\right|=\left|m_i\frac{(\dot{\sigma_e}^2-\sigma_e^2)+m_i(\dot{\sigma_\alpha}^2-\sigma_\alpha^2)}{(\sigma_e^2+m_i\sigma_\alpha^2)(\dot{\sigma_e}^2+m_i\dot{\sigma_\alpha}^2)}\right|\le g^{-\frac{1}{2}}M,\\
&\left|\frac{\zeta_i}{\dot{\zeta}_i}-1\right|=\left|\frac{(\dot{\sigma_e}^2+m_i\dot{\sigma_\alpha}^2)}{(\sigma_e^2+m_i\sigma_\alpha^2)}-1\right|=\left|\frac{(\dot{\sigma_e}^2-\sigma_e^2)+m_i(\dot{\sigma_\alpha}^2-\sigma_\alpha^2)}{(\sigma_e^2+m_i\sigma_\alpha^2)}\right|\le g^{-\frac{1}{2}}M.
\end{aligned}
\tag{3.3.3}
$$

Next, applying the results of (3.2.18), (3.2.19), (3.2.20), (3.2.21), (3.2.24) and (3.2.25), we calculate the terms in $\| \mathrm{E}\{\nabla\psi_i^{(b)}(\boldsymbol{\omega}) - \nabla\psi_i^{(b)}(\boldsymbol{\omega}_0)\}\|$ as follow

$$|\,\mathrm{E}\{l_{i\beta_0\beta_2}(\boldsymbol{\omega})-l_{i\beta_0\beta_2}(\boldsymbol{\omega}_0)\}| = |\,\mathrm{E}\{l_{i\beta_1\beta_2}(\boldsymbol{\omega})-\mathrm{E}\,l_{i\beta_1\beta_2}(\boldsymbol{\omega}_0)\}| = |\,\mathrm{E}\{l_{i\sigma_\alpha^2\beta_2}(\boldsymbol{\omega})-l_{i\sigma_\alpha^2\beta_2}(\boldsymbol{\omega}_0)\}| = 0,$$

$$\begin{aligned}
&|\,\mathrm{E}\{l_{i\beta_0\beta_0}(\boldsymbol{\omega}) - l_{i\beta_0\beta_0}(\boldsymbol{\omega}_0)\}| = |\zeta_i - \dot{\zeta}_i| \le g^{-\frac{1}{2}}M,\\
&|\,\mathrm{E}\{l_{i\beta_0\beta_1}(\boldsymbol{\omega}) - l_{i\beta_0\beta_1}(\boldsymbol{\omega}_0)\}| = |(\zeta_i - \dot{\zeta}_i)\bar{x}_{i.}| \le g^{-\frac{1}{2}}M,\\
&|\,\mathrm{E}\{l_{i\beta_0\sigma_\alpha^2}(\boldsymbol{\omega}) - l_{i\beta_0\sigma_\alpha^2}(\boldsymbol{\omega}_0)\}| = |\zeta_i h_i(\beta_0,\beta_1)| \le g^{-\frac{1}{2}}M,\\
&|\,\mathrm{E}\{l_{i\beta_0\sigma_e^2}(\boldsymbol{\omega}) - l_{i\beta_0\sigma_e^2}(\boldsymbol{\omega}_0)\}| = |m_i^{-1}\zeta_i h_i(\beta_0,\beta_1)| \le g^{-\frac{1}{2}}M,\\
&|\,\mathrm{E}\{l_{i\beta_1\beta_0}(\boldsymbol{\omega}) - l_{i\beta_1\beta_0}(\boldsymbol{\omega}_0)\}| = |(\zeta_i - \dot{\zeta}_i)\bar{x}_{i.}| \le g^{-\frac{1}{2}}M,\\
&|\,\mathrm{E}\{l_{i\beta_1\beta_1}(\boldsymbol{\omega}) - l_{i\beta_1\beta_1}(\boldsymbol{\omega}_0)\}| = |(\zeta_i - \dot{\zeta}_i)\bar{x}_{i.}^2| \le g^{-\frac{1}{2}}M,\\
&|\,\mathrm{E}\{l_{i\beta_1\sigma_\alpha^2}(\boldsymbol{\omega}) - \mathrm{E}\,l_{i\beta_1\sigma_\alpha^2}(\boldsymbol{\omega}_0)\}| = |\zeta_i^2\bar{x}_{i.}h_i(\beta_0,\beta_1) \le g^{-\frac{1}{2}}M,\\
&|\,\mathrm{E}\{l_{i\beta_1\sigma_e^2}(\boldsymbol{\omega}) - \mathrm{E}\,l_{i\beta_1\sigma_e^2}(\boldsymbol{\omega}_0)\}| = |m_i^{-1}\zeta_i^2 h_i(\beta_0,\beta_1)| \le g^{-\frac{1}{2}}M,\\
&|\,\mathrm{E}\{l_{i\sigma_\alpha^2\beta_0}(\boldsymbol{\omega}) - l_{i\sigma_\alpha^2\beta_0}(\boldsymbol{\omega}_0)\}| = |\zeta_i^2 h_i(\beta_0,\beta_1)| \le g^{-\frac{1}{2}}M,\\
&|\,\mathrm{E}\{l_{i\sigma_\alpha^2\beta_1}(\boldsymbol{\omega}) - l_{i\sigma_\alpha^2\beta_1}(\boldsymbol{\omega}_0)\}| = |\zeta_i^2\bar{x}_{i.}h_i(\beta_0,\beta_1)| \le g^{-\frac{1}{2}}M,
\end{aligned}$$

$$\begin{aligned}
&|\,\mathrm{E}\{l_{i\sigma_\alpha^2\sigma_\alpha^2}(\boldsymbol{\omega}) - l_{i\sigma_\alpha^2\sigma_\alpha^2}(\boldsymbol{\omega}_0)\}|\\
&= |\frac{1}{2}\zeta_i^2 - \zeta_i^3\{h_i^2(\beta_0,\beta_1) + \frac{1}{\dot{\zeta}_i}\} + \frac{1}{2}\dot{\zeta}_i^2|\\
&\le |\frac{1}{2}\zeta_i^2 - \frac{\zeta_i^3}{\dot{\zeta}_i} + \frac{1}{2}\dot{\zeta}_i^2| + |\zeta_i^3 h_i^2(\beta_0,\beta_1)|\\
&= |(\zeta_i^2 - \frac{1}{2}\zeta_i^2) - \frac{\zeta_i^3}{\dot{\zeta}_i} + \frac{1}{2}\dot{\zeta}_i^2| + |\zeta_i^3 h_i^2(\beta_0,\beta_1)|\\
&= |\zeta_i^2(1 - \frac{\zeta_i}{\dot{\zeta}_i}) + \frac{1}{2}(\dot{\zeta}_i + \zeta_i)(\dot{\zeta}_i - \zeta_i)| + |\zeta_i^3 h_i^2(\beta_0,\beta_1)|\\
&\le |\zeta_i^2(1 - \frac{\zeta_i}{\dot{\zeta}_i})| + \frac{1}{2}|(\dot{\zeta}_i + \zeta_i)||(\dot{\zeta}_i - \zeta_i)| + |\zeta_i^3 h_i^2(\beta_0,\beta_1)|\\
&\le g^{-\frac{1}{2}}M,
\end{aligned}$$

and

$$|\,\mathrm{E}\{l_{i\sigma_\alpha^2\sigma_e^2}(\boldsymbol{\omega}) - l_{i\sigma_\alpha^2\sigma_e^2}(\boldsymbol{\omega}_0)\}| = m_i^{-1}|\frac{1}{2}\zeta_i^2 - \zeta_i^3\{h_i^2(\beta_0,\beta_1) + \frac{1}{\dot{\zeta}_i}\} + \frac{1}{2}\dot{\zeta}_i^2| \le g^{-\frac{1}{2}}m_i^{-1}M.$$

It follows that

$$\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} g^{-\frac{1}{2}}\sum_{i=1}^{g}|\{\mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega})\right\} - \mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega}_0)\right\}\}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)| = O(g^{-\frac{1}{2}}).$$

The second equation of lemma 15 again requires more care. From (3.2.26), (3.2.27), (3.2.28) and (3.2.29), we have

$$\begin{aligned}
&\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} n^{-\frac{1}{2}}\sum_{i=1}^{g}\sum_{j=1}^{m_i}\left|\left[\mathrm{E}\left\{\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega})\right\}-\mathrm{E}\,\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega}_0)\right](\boldsymbol{\omega}-\boldsymbol{\omega}_0)\right| \\
&\le \sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} n^{-\frac{1}{2}}\sum_{i=1}^{g}\sum_{j=1}^{m_i}\left[\left|\left\{\mathrm{E}\,\frac{l_{ij\beta_2}(\boldsymbol{\omega})}{\partial\boldsymbol{\omega}}-\mathrm{E}\,\frac{l_{ij\beta_2}(\boldsymbol{\omega}_0)}{\partial\boldsymbol{\omega}}\right\}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)\right|\right. \\
&\qquad\left.+\left|\left\{\mathrm{E}\,\frac{l_{ij\sigma_e^2}(\boldsymbol{\omega})}{\partial\boldsymbol{\omega}}-\mathrm{E}\,\frac{l_{ij\sigma_e^2}(\boldsymbol{\omega}_0)}{\partial\boldsymbol{\omega}}\right\}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)\right|\right] \\
&\le \sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} n^{-\frac{1}{2}}\sum_{i=1}^{g}\sum_{j=1}^{m_i}\Big[\left|\left\{\mathrm{E}\,l_{ij\beta_2\beta_2}(\boldsymbol{\omega})-\mathrm{E}\,l_{ij\beta_2\beta_2}(\boldsymbol{\omega}_0)\right\}(\beta_2-\dot\beta_2)\right|+\left|\mathrm{E}\,l_{ij\beta_2\sigma_e^2}(\boldsymbol{\omega})(\sigma_e^2-\dot{\sigma_e}^2)\right| \\
&\qquad+\left|\mathrm{E}\,l_{ij\sigma_e^2\beta_0}(\boldsymbol{\omega})(\beta_0-\dot\beta_0)\right|+\left|\mathrm{E}\,l_{ij\sigma_e^2\beta_1}(\boldsymbol{\omega})(\beta_1-\dot\beta_1)\right|+\left|\mathrm{E}\,l_{ij\sigma_e^2\sigma_\alpha^2}(\boldsymbol{\omega})(\sigma_\alpha^2-\dot{\sigma_\alpha}^2)\right| \\
&\qquad+\left|\mathrm{E}\,l_{ij\sigma_e^2\beta_2}(\boldsymbol{\omega})(\beta_2-\dot\beta_2)\right|+\left|\left\{\mathrm{E}\,l_{ij\sigma_e^2\sigma_e^2}(\boldsymbol{\omega})-\mathrm{E}\,l_{ij\sigma_e^2\sigma_e^2}(\boldsymbol{\omega}_0)\right\}(\sigma_e^2-\dot{\sigma_e}^2)\right|\Big].
\end{aligned}$$

Note that some terms such as $\mathrm{E}\{l_{ij\beta_1\sigma_e^2}(\boldsymbol{\omega}_0)\}=0$ have already been set to zero in the above upper bound. Since $|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M$, as $g,m\to\infty$, then we can show that $|\beta_0-\dot\beta_0|\le g^{-\frac{1}{2}}M$, $|\beta_1-\dot\beta_1|\le g^{-\frac{1}{2}}M$, $|\sigma_\alpha^2-\dot{\sigma_\alpha}^2|\le g^{-\frac{1}{2}}M$, $|\beta_2-\dot\beta_2|\le n^{-\frac{1}{2}}M$, and $|\sigma_e^2-\dot{\sigma_e}^2|\le n^{-\frac{1}{2}}M$. Thus combining the results of (3.2.26), (3.2.27), (3.2.28) and (3.2.29), we get

$$\begin{aligned}
&|\{\mathrm{E}\,l_{ij\beta_2\beta_2}(\boldsymbol{\omega})-\mathrm{E}\,l_{ij\beta_2\beta_2}(\boldsymbol{\omega}_0)\}(\beta_2-\dot\beta_2)|=\left|\frac{\sigma_e^2-\dot{\sigma_e}^2}{\sigma_e^2\dot{\sigma_e}^2}S_{ijw}^x(\beta_2-\dot\beta_2)\right|\le n^{-1}M, \\
&|\mathrm{E}\,l_{ij\beta_2\sigma_e^2}(\boldsymbol{\omega})(\sigma_e^2-\dot{\sigma_e}^2)|=\left|-\frac{1}{\sigma_e^4}(\dot\beta_2-\beta_2)(\sigma_e^2-\dot{\sigma_e}^2)\right|\le n^{-1}M, \\
&|\mathrm{E}\,l_{ij\sigma_e^2\beta_0}(\boldsymbol{\omega})(\beta_0-\dot\beta_0)|=\left|\frac{-1}{(\sigma_e^2+m_i\sigma_\alpha^2)^2}[(\dot\beta_0-\beta_0)+\bar x_{i.}(\dot\beta_1-\beta_1)](\dot\beta_0-\beta_0)\right|\le g^{-1}m_i^{-2}M, \\
&|\mathrm{E}\,l_{ij\sigma_e^2\beta_1}(\boldsymbol{\omega})(\beta_1-\dot\beta_1)|=\left|\frac{-\bar x_{i.}}{(\sigma_e^2+m_i\sigma_\alpha^2)^2}[(\dot\beta_0-\beta_0)+\bar x_{i.}(\dot\beta_1-\beta_1)](\dot\beta_1-\beta_1)\right|\le g^{-1}m_i^{-2}M, \\
&|\mathrm{E}\,l_{ij\sigma_e^2\beta_2}(\boldsymbol{\omega})(\beta_2-\dot\beta_2)|=\left|\frac{1}{\sigma_e^4}S_{ijw}^x(\dot\beta_2-\beta_2)^2\right|\le n^{-1}M,
\end{aligned}$$

$$\begin{aligned}
|\{\mathrm{E}\,l_{ij\sigma_e^2\sigma_\alpha^2}(\boldsymbol{\omega})-\mathrm{E}\,l_{ij\sigma_e^2\sigma_\alpha^2}(\boldsymbol{\omega}_0)\}(\sigma_\alpha^2-\dot{\sigma_\alpha}^2)| &\le m_i^{-2}\left|\frac{1}{2}\zeta_i^2-\zeta_i^3\{h_i^2(\beta_0,\beta_1)+\frac{1}{\dot\zeta_i}\}+\frac{1}{2}\dot\zeta_i^2\right|\left|(\sigma_\alpha^2-\dot{\sigma_\alpha}^2)\right| \\
&\le g^{-1}m_i^{-2}M,
\end{aligned}$$

and

$$
\begin{aligned}
&|\,\mathrm{E}\{l_{ij\sigma_e^2\sigma_e^2}(\boldsymbol{\omega}) - \mathrm{E}\, l_{ij\sigma_e^2\sigma_e^2}(\boldsymbol{\omega}_0)\}(\sigma_e^2 - \dot{\sigma_e}^2)| \\
&= |\left[\frac{1}{2}m_i^{-3}\zeta^2 - m_i^{-3}\zeta^3\left\{h_i^2(\beta_0,\beta_1) + \frac{1}{\dot{\zeta}}\right\} + \frac{n-g}{2n\sigma_e^4} - \frac{1}{\sigma_e^6}\left\{S_{ijw}^x(\dot{\beta}_2 - \beta_2)^2 - \frac{n-g}{n}\dot{\sigma_e}^2\right\}\right. \\
&\quad \left. + \frac{1}{2}m_i^{-3}\dot{\zeta}^2 + \frac{n-g}{2n\dot{\sigma_e}^4}\right](\sigma_e^2 - \dot{\sigma_e}^2)| \\
&\leq |m_i^{-3}\left[\frac{1}{2}\zeta^2 - m_i^{-3}\zeta^3\left\{h_i^2(\beta_0,\beta_1) + \frac{1}{\dot{\zeta}}\right\} + \frac{1}{2}m_i^{-3}\dot{\zeta}^2\right](\sigma_e^2 - \dot{\sigma_e}^2)| \\
&\quad + |\left[\frac{n-g}{2n\sigma_e^4} - \frac{1}{\sigma_e^6}\left\{S_{ijw}^x(\dot{\beta}_2 - \beta_2)^2 - \frac{n-g}{n}\dot{\sigma_e}^2\right\} + \frac{n-g}{2g\dot{\sigma_e}^4}\right](\sigma_e^2 - \dot{\sigma_e}^2)| \\
&\leq g^{-1}m_i^{-\frac{7}{2}}M + |\left[\frac{n-g}{2n\sigma_e^4} - \frac{1}{\sigma_e^6}\left\{S_{ijw}^x(\dot{\beta}_2 - \beta_2)^2 - \frac{n-g}{n}\dot{\sigma_e}^2\right\} + \frac{n-g}{2n\dot{\sigma_e}^4}\right](\sigma_e^2 - \dot{\sigma_e}^2)| \\
&\leq g^{-1}m_i^{-\frac{7}{2}}M + |\left(\frac{n-g}{2n\sigma_e^4} - \frac{1}{\sigma_e^6}\frac{n-g}{n}\dot{\sigma_e}^2 + \frac{n-g}{2n\dot{\sigma_e}^4}\right)(\sigma_e^2 - \dot{\sigma_e}^2)| + |\frac{1}{\sigma_e^6}S_{ijw}^x(\dot{\beta}_2 - \beta_2)^2(\sigma_e^2 - \dot{\sigma_e}^2)| \\
&\leq g^{-1}m_i^{-\frac{7}{2}}M + g^{-\frac{3}{2}}m_i^{-\frac{3}{2}}M + |\left(\frac{n-g}{2g\sigma_e^4} - \frac{1}{\sigma_e^6}\frac{n-g}{n}\dot{\sigma_e}^2 + \frac{n-g}{2n\dot{\sigma_e}^4}\right)(\sigma_e^2 - \dot{\sigma_e}^2)| \\
&= g^{-1}m_i^{-\frac{7}{2}}M + g^{-\frac{3}{2}}m_i^{-\frac{3}{2}}M + \frac{n-g}{n}|\left(\frac{1}{\sigma_e^4} - \frac{1}{2\sigma_e^4} - \frac{1}{\sigma_e^6}\dot{\sigma_e}^2 + \frac{1}{2\dot{\sigma_e}^4}\right)(\sigma_e^2 - \dot{\sigma_e}^2)| \\
&\leq g^{-1}m_i^{-\frac{7}{2}}M + g^{-\frac{3}{2}}m_i^{-\frac{3}{2}}M + \frac{n-g}{n}|\left(\frac{1}{\sigma_e^4} - \frac{1}{2\sigma_e^4} - \frac{1}{\sigma_e^6}\dot{\sigma_e}^2 + \frac{1}{2\dot{\sigma_e}^4}\right)(\sigma_e^2 - \dot{\sigma_e}^2)| \\
&\leq g^{-1}m_i^{-\frac{7}{2}}M + g^{-\frac{3}{2}}m_i^{-\frac{3}{2}}M + \frac{n-g}{n}|\frac{1}{\sigma_e^4}\left(1 - \frac{\dot{\sigma_e}^2}{\sigma_e^2}\right) + \frac{1}{2}\left(\frac{1}{\dot{\sigma_e}^4} - \frac{1}{\sigma_e^4}\right)||(\sigma_e^2 - \dot{\sigma_e}^2)| \\
&\leq g^{-1}m_i^{-\frac{7}{2}}M + g^{-\frac{3}{2}}m_i^{-\frac{3}{2}}M + n^{-1}M
\end{aligned}
$$

Therefore, we obtain

$$
\begin{aligned}
&\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\leq M} n^{-\frac{1}{2}}\sum_{i=1}^{g}\sum_{j=1}^{m_i}|\left[\mathrm{E}\left\{\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega})\right\} - \mathrm{E}\,\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega}_0)\right](\boldsymbol{\omega} - \boldsymbol{\omega}_0)| \\
&= \min_{1\leq i\leq g}\{O(n^{-\frac{1}{2}}), O(n^{\frac{1}{2}}g^{-1}m_i^{-2})\} \\
&= o(1).
\end{aligned}
$$

□

Lemma 16. *Suppose Condition B holds. Then, as $g \to \infty$ and $m_L \to \infty$,*

$$
\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\leq M} |\mathbf{K}^{-\frac{1}{2}}\left[\psi(\boldsymbol{\omega}) - \psi(\boldsymbol{\omega}_0) - \mathrm{E}\left\{\psi(\boldsymbol{\omega}) - \psi(\boldsymbol{\omega}_0)\right\}\right]| = o_p(1).
$$

Proof. Decompose the cube $\mathcal{B} = \{\boldsymbol{\omega} : |\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega} - \boldsymbol{\omega}_0)| \leq M\}$ into smaller cubes $\mathcal{C} = \{\mathcal{C}(t_k)\}$, where $M > 0$ is a constant and $\{\mathcal{C}(t_k)\}$ is a cube containing $\boldsymbol{\omega} \in \mathcal{C}(t_k) = \{\boldsymbol{\omega} : |\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega} - \boldsymbol{\omega}_0)| \leq Mg^{-\frac{1}{4}}\}$. Let $N = g^{\frac{1}{4}}$.We first show that lemma 16holds over the $\mathcal{C}$ of fixed $t_k =$

$(t_{k1}, t_{k2}, t_{k3}, t_{k4}, t_{k5})^T$. Using Chebyshev's inequality, for any $\eta > 0$, we have

$$
\begin{aligned}
&\sum_{k=1}^{N} \Pr\left(|\mathbf{K}^{-\frac{1}{2}}\left[\psi(t_k) - \psi(\boldsymbol{\omega}_0) - \mathrm{E}\left\{\psi(t_k) - \psi(\boldsymbol{\omega}_0)\right\}\right]| > \eta\right) \\
&\leq \eta^{-2} \sum_{k=1}^{N} \mathrm{E}\,|\mathbf{K}^{-\frac{1}{2}}\left[\psi(t_k) - \psi(\boldsymbol{\omega}_0) - \mathrm{E}\left\{\psi(t_k) - \psi(\boldsymbol{\omega}_0)\right\}\right]|^2 \\
&= \eta^{-2} \sum_{k=1}^{N} \mathrm{E}\,|g^{-\frac{1}{2}}\left[\psi^{(b)}(t_k) - \psi^{(b)}(\boldsymbol{\omega}_0) - \mathrm{E}\left\{\psi^{(b)}(t_k) - \psi^{(b)}(\boldsymbol{\omega}_0)\right\}\right]|^2 \\
&\quad + \eta^{-2} \sum_{k=1}^{N} \mathrm{E}\,|n^{-\frac{1}{2}}\left[\psi^{(w)}(t_k) - \psi^{(w)}(\boldsymbol{\omega}_0) - \mathrm{E}\left\{\psi^{(w)}(t_k) - \psi^{(w)}(\boldsymbol{\omega}_0)\right\}\right]|^2 \\
&= \eta^{-2} g^{-1} \sum_{k=1}^{N} \mathrm{E}\left[|\psi^{(b)}(t_k) - \psi^{(b)}(\boldsymbol{\omega}_0) - \mathrm{E}\left\{\psi^{(b)}(t_k) - \psi^{(b)}(\boldsymbol{\omega}_0)\right\}|^2\right] \\
&\quad + \eta^{-2} n^{-1} \sum_{k=1}^{N} \mathrm{E}\left[|\psi^{(w)}(t_k) - \psi^{(w)}(\boldsymbol{\omega}_0) - \mathrm{E}\left\{\psi^{(w)}(t_k) - \psi^{(w)}(\boldsymbol{\omega}_0)\right\}|^2\right].
\end{aligned} \tag{3.3.4}
$$

For the first term of (3.3.4), we can use the result of lemma 13 and a Taylor expansion to obtain

$$
\begin{aligned}
&\eta^{-2} g^{-1} \sum_{k=1}^{N} \mathrm{E}\left[|\psi^{(b)}(t_k) - \psi^{(b)}(\boldsymbol{\omega}_0) - \mathrm{E}\left\{\psi^{(b)}(t_k) - \psi^{(b)}(\boldsymbol{\omega}_0)\right\}|^2\right] \\
&\leq \eta^{-2} g^{-1} \sum_{k=1}^{N} \mathrm{E}\left\{|\psi^{(b)}(t_k) - \psi^{(b)}(\boldsymbol{\omega}_0)|^2\right\} \\
&= \eta^{-2} g^{-1} \sum_{k=1}^{N} \mathrm{E}\left[|\sum_{i=1}^{g}\{\psi_i^{(b)}(t_k) - \psi_i^{(b)}(\boldsymbol{\omega}_0)\right\}|^2] \\
&\leq \eta^{-2} g^{-1} \sum_{k=1}^{N} \sum_{i=1}^{g} \mathrm{E}\left\{|\psi_i^{(b)}(t_k) - \psi_i^{(b)}(\boldsymbol{\omega}_0)|^2\right\} \\
&= \eta^{-2} g^{-1} \sum_{k=1}^{N} \sum_{i=1}^{g} \mathrm{E}\left\{|\nabla\psi_i^{(b)}(\tilde{\boldsymbol{\omega}})(t_k - \boldsymbol{\omega}_0)|^2\right\},
\end{aligned}
$$

where $\tilde{\boldsymbol{\omega}}$ is between t_k and $\boldsymbol{\omega}_0$. Because $\|\mathbf{K}^{-\frac{1}{2}}\| \leq g^{-\frac{1}{2}} L$ for a constant $L > 0$, then it follows that

$$
\begin{aligned}
\eta^{-2} g^{-1} \sum_{k=1}^{N} \sum_{i=1}^{g} \mathrm{E}\left\{|\nabla\psi_i^{(b)}(\tilde{\boldsymbol{\omega}})(t_k - \boldsymbol{\omega}_0)|^2\right\} &\leq \eta^{-2} g^{-1} \sum_{k=1}^{N} \sum_{i=1}^{g} \mathrm{E}\left\{\|\nabla\psi_i^{(b)}(\tilde{\boldsymbol{\omega}})\|^2 \|\mathbf{K}^{-\frac{1}{2}}\|^2 |\mathbf{K}^{\frac{1}{2}}(t_k - \boldsymbol{\omega}_0)|^2\right\} \\
&\leq \eta^{-2} g^{-2} M^2 L^2 \sum_{k=1}^{N} \sum_{i=1}^{g} \mathrm{E}\,\|\nabla\psi_i^{(b)}(\tilde{\boldsymbol{\omega}})\|^2 \\
&= O_p(g^{-\frac{3}{4}}) = o_p(1).
\end{aligned}
$$

Similarly, for the second term of (3.3.4), we apply the result in lemma 13 and a Taylor expansion to obtain

$$
\begin{aligned}
&\eta^{-2}n^{-1}\sum_{k=1}^{N}\mathrm{E}\left[|\psi^{(w)}(t_k)-\psi^{(w)}(\boldsymbol{\omega}_0)-\mathrm{E}\left\{\psi^{(w)}(t_k)-\psi^{(w)}(\boldsymbol{\omega}_0)\right\}|^2\right]\\
&\leq \eta^{-2}n^{-1}\sum_{k=1}^{N}\mathrm{E}\left\{|\psi^{(w)}(t_k)-\psi^{(w)}(\boldsymbol{\omega}_0)|^2\right\}\\
&= \eta^{-2}n^{-1}\sum_{k=1}^{N}\mathrm{E}\left[|\sum_{i=1}^{g}\sum_{j=1}^{m_i}\left\{\psi_{ij}^{(w)}(t_k)-\psi_{ij}^{(w)}(\boldsymbol{\omega}_0)\right\}|^2\right]\\
&\leq \eta^{-2}n^{-1}\sum_{k=1}^{N}\sum_{i=1}^{g}\sum_{j=1}^{m_i}\mathrm{E}\left\{|\psi_{ij}^{(w)}(t_k)-\psi_{ij}^{(w)}(\boldsymbol{\omega}_0)|^2\right\}\\
&= \eta^{-2}n^{-1}\sum_{k=1}^{N}\sum_{i=1}^{g}\sum_{j=1}^{m_i}\mathrm{E}\left\{|\nabla\psi_{ij}^{(w)}(\tilde{\boldsymbol{\omega}})(t_k-\boldsymbol{\omega}_0)|^2\right\}\\
&\leq \eta^{-2}n^{-1}\sum_{k=1}^{N}\sum_{i=1}^{g}\sum_{j=1}^{m_i}\mathrm{E}\left\{\|\nabla\psi_{ij}^{(w)}(\tilde{\boldsymbol{\omega}})\|^2\|\mathbf{K}^{-\frac{1}{2}}\|^2|\mathbf{K}^{\frac{1}{2}}(t_k-\boldsymbol{\omega}_0)|^2\right\}\\
&\leq \eta^{-2}g^{-1}n^{-1}M^2L^2\sum_{k=1}^{N}\sum_{i=1}^{g}\sum_{j=1}^{m_i}\mathrm{E}[\|\nabla\psi_{ij}^{(w)}(\tilde{\boldsymbol{\omega}})\|^2\\
&= O_p(g^{-\frac{3}{4}}).
\end{aligned}
$$

Next, we show that the difference between taking the supermum over a fine grid of points and over $\mathcal{B}$ is small. Using a Taylor expansion, we have

$$
\begin{aligned}
&\max_{1\leq k\leq N}\sup_{\boldsymbol{\omega}\in\mathcal{C}(t_k)}|\mathbf{K}^{-\frac{1}{2}}\left[\psi(\boldsymbol{\omega})-\psi(t_k)-\mathrm{E}\left\{\psi(\boldsymbol{\omega})-\psi(t_k)\right\}\right]|\\
\leq{}& \max_{1\leq k\leq N}\sup_{\boldsymbol{\omega}\in\mathcal{C}(t_k)} g^{-\frac{1}{2}}|\psi^{(b)}(\boldsymbol{\omega})-\psi^{(b)}(t_k)-\mathrm{E}\left\{\psi^{(b)}(\boldsymbol{\omega})-\psi^{(b)}(t_k)\right\}|\\
&+\max_{1\leq k\leq N}\sup_{\boldsymbol{\omega}\in\mathcal{C}(t_k)} n^{-\frac{1}{2}}|\psi^{(w)}(\boldsymbol{\omega})-\psi^{(w)}(t_k)-\mathrm{E}\left\{\psi^{(w)}(\boldsymbol{\omega})-\psi^{(w)}(t_k)\right\}|\\
={}& \max_{1\leq k\leq N}\sup_{\boldsymbol{\omega}\in\mathcal{C}(t_k)} g^{-\frac{1}{2}}|\sum_{i=1}^{g}\left[\psi_i^{(b)}(\boldsymbol{\omega})-\psi_i^{(b)}(t_k)-\mathrm{E}\left\{\psi_i^{(b)}(\boldsymbol{\omega})-\psi_i^{(b)}(t_k)\right\}\right]|\\
&+\max_{1\leq k\leq N}\sup_{\boldsymbol{\omega}\in\mathcal{C}(t_k)} n^{-\frac{1}{2}}|\sum_{i=1}^{g}\sum_{j=1}^{m_i}\left[\psi_{ij}^{(w)}(\boldsymbol{\omega})-\psi_{ij}^{(w)}(t_k)-\mathrm{E}\left\{\psi_{ij}^{(w)}(\boldsymbol{\omega})-\psi_{ij}^{(w)}(t_k)\right\}\right]|\\
\leq{}& \max_{1\leq k\leq N}\sup_{\boldsymbol{\omega}\in\mathcal{C}(t_k)} g^{-\frac{1}{2}}\sum_{i=1}^{g}|\psi_i^{(b)}(\boldsymbol{\omega})-\psi_i^{(b)}(t_k)-\mathrm{E}\left\{\psi_i^{(b)}(\boldsymbol{\omega})-\psi_i^{(b)}(t_k)\right\}|\\
&+\max_{1\leq k\leq N}\sup_{\boldsymbol{\omega}\in\mathcal{C}(t_k)} n^{-\frac{1}{2}}\sum_{i=1}^{g}\sum_{j=1}^{m_i}|\psi_{ij}^{(w)}(\boldsymbol{\omega})-\psi_{ij}^{(w)}(t_k)-\mathrm{E}\left\{\psi_{ij}^{(w)}(\boldsymbol{\omega})-\psi_{ij}^{(w)}(t_k)\right\}|\\
\leq{}& \max_{1\leq k\leq N}\sup_{\boldsymbol{\omega}\in\mathcal{C}(t_k)} g^{-\frac{1}{2}}\sum_{i=1}^{g}|\left\{\nabla\psi_i^{(b)}(\tilde{\boldsymbol{\omega}}_1)-\mathrm{E}\,\nabla\psi_i^{(b)}(\boldsymbol{\omega}_1^*)\right\}(\boldsymbol{\omega}-t_k)|\\
&+\max_{1\leq k\leq N}\sup_{\boldsymbol{\omega}\in\mathcal{C}(t_k)} n^{-\frac{1}{2}}\sum_{i=1}^{g}\sum_{j=1}^{m_i}|\left\{\nabla\psi_{ij}^{(w)}(\tilde{\boldsymbol{\omega}}_2)-\mathrm{E}\,\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega}_2^*)\right\}(\boldsymbol{\omega}-t_k)|
\end{aligned}
\tag{3.3.5}
$$

where $\tilde{\boldsymbol{\omega}}_1, \tilde{\boldsymbol{\omega}}_2, \boldsymbol{\omega}_1^*$ and $\boldsymbol{\omega}_2^*$ are between t_k and $\boldsymbol{\omega}$. The required result follows from lemma 14. □

Lemma 17. *Suppose Condition B holds. Then, as $g \to \infty$ and $m_L \to \infty$,*

$$\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\leq M} |\mathbf{K}^{-\frac{1}{2}}[\mathrm{E}\left\{\psi(\boldsymbol{\omega})-\psi(\boldsymbol{\omega}_0)\right\}-\mathrm{E}\left\{\nabla\psi(\boldsymbol{\omega}_0)\right\}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)]|=o_p(1).$$

Proof. Write

$$\begin{aligned}
&\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\leq M} |\mathbf{K}^{-\frac{1}{2}}\left[\mathrm{E}\left\{\psi(\boldsymbol{\omega})-\psi(\boldsymbol{\omega}_0)\right\}-\mathrm{E}\left\{\nabla\psi(\boldsymbol{\omega}_0)\right\}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)\right]| \\
\leq &\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\leq M} |g^{-\frac{1}{2}}\left[\mathrm{E}\left\{\psi^{(b)}(\boldsymbol{\omega})-\psi^{(b)}(\boldsymbol{\omega}_0)\right\}-\mathrm{E}\left\{\nabla\psi^{(b)}(\boldsymbol{\omega}_0)\right\}(\boldsymbol{\omega}^{(b)}-\boldsymbol{\omega}_0^{(b)})\right]| \\
&+\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\leq M} |n^{-\frac{1}{2}}\left[\mathrm{E}\left\{\psi^{(w)}(\boldsymbol{\omega})-\psi^{(w)}(\boldsymbol{\omega}_0)\right\}-\mathrm{E}\left\{\nabla\psi^{(w)}(\boldsymbol{\omega}_0)\right\}(\boldsymbol{\omega}^{(w)}-\boldsymbol{\omega}_0^{(w)})\right]| \\
= &\sup_{|\mathbf{K}^{-\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\leq M} g^{-\frac{1}{2}}|\sum_{i=1}^{g}\left[\mathrm{E}\left\{\psi_i^{(b)}(\boldsymbol{\omega})-\psi_i^{(b)}(\boldsymbol{\omega}_0)\right\}-\mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega}_0)\right\}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)\right]| \\
&+\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\leq M} n^{-\frac{1}{2}}|\sum_{i=1}^{g}\sum_{j=1}^{m_i}\left[\mathrm{E}\left\{\psi_{ij}^{(w)}(\boldsymbol{\omega})-\psi_{ij}^{(w)}(\boldsymbol{\omega}_0)\right\}-\mathrm{E}\left\{\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega}_0)\right\}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)\right]| \\
\leq &\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\leq M} g^{-\frac{1}{2}}\sum_{i=1}^{g}|\,\mathrm{E}\left\{\psi_i^{(b)}(\boldsymbol{\omega})-\psi_i^{(b)}(\boldsymbol{\omega}_0)\right\}-\mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega}_0)\right\}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)| \\
&+\sup_{|(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\leq M} n^{-\frac{1}{2}}\sum_{i=1}^{g}\sum_{j=1}^{m_i}|\,\mathrm{E}\left\{\psi_{ij}^{(w)}(\boldsymbol{\omega})-\psi_{ij}^{(w)}(\boldsymbol{\omega}_0)\right\}-\mathrm{E}\left\{\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega}_0)\right\}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)| \\
= &\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\leq M} g^{-\frac{1}{2}}\sum_{i=1}^{g}|\left[\mathrm{E}\left\{\nabla\psi_i^{(b)}(\tilde{\boldsymbol{\omega}}_1)\right\}-\mathrm{E}\left\{\nabla\psi_i^{(b)}(\boldsymbol{\omega}_0)\right\}\right](\boldsymbol{\omega}-\boldsymbol{\omega}_0)| \\
&+\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\leq M} n^{-\frac{1}{2}}\sum_{i=1}^{g}\sum_{j=1}^{m_i}|\left[\mathrm{E}\left\{\nabla\psi_{ij}^{(w)}(\tilde{\boldsymbol{\omega}}_2)\right\}-\mathrm{E}\left\{\nabla\psi_{ij}^{(w)}(\boldsymbol{\omega}_0)\right\}\right](\boldsymbol{\omega}-\boldsymbol{\omega}_0)|,
\end{aligned}$$

where $\tilde{\boldsymbol{\omega}}_1$ and $\tilde{\boldsymbol{\omega}}_2$ are between $\boldsymbol{\omega}$ and $\boldsymbol{\omega}_0$. Then the required result follows from applying lemma 15. □

3.3.1 Proof of the main theorem

Proof. We write

$$\begin{aligned}
\mathbf{K}^{-\frac{1}{2}}\psi(\boldsymbol{\omega}) = \mathbf{K}^{-\frac{1}{2}}\left[\psi(\boldsymbol{\omega})-\psi(\boldsymbol{\omega}_0)-\mathrm{E}\left\{\psi(\boldsymbol{\omega})-\psi(\boldsymbol{\omega}_0)\right\}\right]&+\mathbf{K}^{-\frac{1}{2}}\left\{\psi(\boldsymbol{\omega}_0)-\xi(\boldsymbol{\omega}_0)\right\} \\
+\mathbf{K}^{-\frac{1}{2}}\left[\mathrm{E}\left\{\psi(\boldsymbol{\omega})-\psi(\boldsymbol{\omega}_0)-\nabla\psi(\boldsymbol{\omega}_0)(\boldsymbol{\omega}-\boldsymbol{\omega}_0)\right\}\right]& \\
+\mathbf{K}^{-\frac{1}{2}}\left[\xi(\boldsymbol{\omega}_0)+\mathrm{E}\left\{\nabla\psi(\boldsymbol{\omega}_0)\right\}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)\right].&
\end{aligned}$$

From (3.2.9), we have have

$$\mathbf{K}^{-\frac{1}{2}}\left\{\psi(\boldsymbol{\omega}_0)-\xi(\boldsymbol{\omega}_0)\right\}=o_p(1).$$

Also, from lemma 16 and lemma 17, we have

$$\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\leq M}|\mathbf{K}^{-\frac{1}{2}}[\psi(\boldsymbol{\omega})-\psi(\boldsymbol{\omega}_0)-\mathrm{E}\{\psi(\boldsymbol{\omega})-\psi(\boldsymbol{\omega}_0)\}]|=o_p(1)\qquad\text{and}$$
$$\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\leq M}|\mathbf{K}^{-\frac{1}{2}}[\mathrm{E}\{\psi(\boldsymbol{\omega})-\psi(\boldsymbol{\omega}_0)\}-\mathrm{E}\{\nabla\psi(\boldsymbol{\omega}_0)\}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)]|=o_p(1).$$

On combining above equations, we can see that, uniformly on $\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)\leq M$,

$$\mathbf{K}^{-\frac{1}{2}}\psi(\boldsymbol{\omega})=\mathbf{K}^{-\frac{1}{2}}\xi(\boldsymbol{\omega}_0)+\mathbf{B}_n\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)+o_p(1), \tag{3.3.6}$$

where $\mathbf{B}_n=-\mathbf{K}^{-\frac{1}{2}}\,\mathrm{E}\{\nabla\psi(\boldsymbol{\omega}_0)\}\mathbf{K}^{-\frac{1}{2}}$. The right-hand side of (3.3.6) is mainly determined by the negative of the quadratic form of $\mathbf{B}_n$ and is negative for M sufficiently large. Therefore, applying to Result 6.3.4 of Ortega and Rheinboldt (1973), we conclude that a solution to the estimating equations exists in probability and satisfies $|\mathbf{K}^{-\frac{1}{2}}(\hat{\boldsymbol{\omega}}-\boldsymbol{\omega}_0)|=O_p(1)$. The asymptotic representation follows from lemma 16and lemma 17. Moreover, Lemma 12 implies that the Lyapounov condition holds and hence the asymptotic normality result follows. □

3.4 Adjusted Maximum Likelihood (AML) Estimating Equations

In this section, we derive the asymptotic properties of REML estimates in nested error regression model for the unbalanced data setting. Again if we use same approach as in Chapter 2, then it will result in excessively tedious calculations along with further restrictive conditions that need to be imposed. This is because we used a balanced structure in Chapter 2, where t $m_i=m$ for $i=1,\ldots,g$. However, in this chapter, we consider the general data case (unbalanced data), which does not allow us to exploit the convenience of a balanced structure. For example, we cannot obtain very elegant results like (2.4.9) in unbalanced data. Thus, we require an alternative approach to obtain asymptotic properties of REML estimators in the nested error regression model. We begin by constructing an adjusted maximum likelihood (AML) in the proceeding section.

Let $\mathbf{X}=[\mathbf{1}_n,\bar{\mathbf{x}},\mathbf{x}-\bar{\mathbf{x}}]$, where $\bar{\mathbf{x}}^T=[\mathbf{1}_{m_1}^T\bar{x}_1,\ldots,\mathbf{1}_{m_g}^T\bar{x}_g]$ and $\mathbf{x}^T=[\mathbf{x}_1^T,\ldots,\mathbf{x}_g^T]$. Then we rewrite (3.1.3) in general matrix from as from

$$\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}_1\boldsymbol{\alpha}+\mathbf{Z}_2\mathbf{e}. \tag{3.4.1}$$

In this case, $\mathbf{Z}_1 = \mathrm{diag}\{\mathbf{1}_{m_1}, \ldots, \mathbf{1}_{m_g}\}$ and $\mathbf{Z}_2 = \mathbf{I}_n$. We can also rewrite (3.1.4) in general form as

$$l(\mathbf{y}; \boldsymbol{\beta}, \boldsymbol{\theta}) = -\frac{1}{2} n \log 2\pi - \frac{1}{2} \log |\mathbf{V}(\boldsymbol{\theta})| - \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{V}^{-1}(\boldsymbol{\theta})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \tag{3.4.2}$$

Therefore, we obtain the adjusted log-likelihood function from (3.4.2) after adding $\log \|\mathbf{X}^T \mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}\|$, resulting in

$$l_A(\mathbf{y}; \boldsymbol{\beta}, \boldsymbol{\theta}) = -\frac{1}{2} n \log 2\pi - \frac{1}{2} \log |\mathbf{V}(\boldsymbol{\theta})| - \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{V}^{-1}(\boldsymbol{\theta})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) - \frac{1}{2} \log \|\mathbf{X}^T \mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}\|. \tag{3.4.3}$$

We denote the adjusted maximum likelihood estimating equation as $\psi_A(\boldsymbol{\omega}) = 0$, where $\psi_A(\boldsymbol{\omega}) = [l_{A\beta_0}(\boldsymbol{\omega}), l_{A\beta_1}(\boldsymbol{\omega}), l_{A\sigma_\alpha^2}(\boldsymbol{\omega}), l_{A\beta_2}(\boldsymbol{\omega}), l_{A\sigma_e^2}(\boldsymbol{\omega})]^T$.

To obtain the asymptotic properties of the AML estimator $\hat{\boldsymbol{\omega}}_A$, we require one more lemma.

Lemma 18. *Suppose Condition B holds. As $g, m_L \to \infty$,*

$$\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega} - \boldsymbol{\omega}_0)| \le M} g^{-\frac{1}{2}} \operatorname{tr}[\{\mathbf{X}^T \mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}\}^{-1} \mathbf{X}^T \mathbf{V}^{-1}(\boldsymbol{\theta}) \mathbf{Z}_1 \mathbf{Z}_1^T \mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}] = o_p(1), \quad \textit{and}$$
$$\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega} - \boldsymbol{\omega}_0)| \le M} n^{-\frac{1}{2}} \operatorname{tr}[\{\mathbf{X}^T \mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}\}^{-1} \mathbf{X}^T \mathbf{V}^{-1}(\boldsymbol{\theta}) \mathbf{Z}_2 \mathbf{Z}_2^T \mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}] = o_p(1).$$

Proof. We calculate

$$\lim_{g, m_L \to \infty} \sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega} - \boldsymbol{\omega}_0)| \le M} \operatorname{tr}[\{\mathbf{X}^T \mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}\}^{-1} \mathbf{X}^T \mathbf{V}^{-1}(\boldsymbol{\theta}) \mathbf{Z}_1 \mathbf{Z}_1^T \mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}] = \frac{2}{\sigma_\alpha^2}$$
$$\lim_{g, m_L \to \infty} \sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega} - \boldsymbol{\omega}_0)| \le M} \operatorname{tr}[\{\mathbf{X}^T \mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}\}^{-1} \mathbf{X}^T \mathbf{V}^{-1}(\boldsymbol{\theta}) \mathbf{Z}_2 \mathbf{Z}_2^T \mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}] = \frac{1}{\sigma_e^2}.$$

Thus, we get the following result. □

Theorem 4. *Suppose Conditions B and C hold. Then there exists a solution $\hat{\boldsymbol{\omega}}_A$ to the adjusted estimating equations $\psi_A(\boldsymbol{\omega}) = 0$, satisfying $|\mathbf{K}^{\frac{1}{2}}(\hat{\boldsymbol{\omega}}_A - \boldsymbol{\omega}_0)| = O_p(1)$, as $g \to \infty$ and $m_L \to \infty$, where $\mathbf{K} = \mathrm{diag}[g\mathbf{I}_{(b)}, n\mathbf{I}_{(w)}]$. Moreover,*

$$\mathbf{K}^{\frac{1}{2}}(\hat{\boldsymbol{\omega}}_A - \boldsymbol{\omega}_0) = \mathbf{B}^{-1}\mathbf{K}^{-\frac{1}{2}}\xi(\boldsymbol{\theta}_0) + o_p(1), \tag{3.4.4}$$

whence

$$\mathbf{K}^{\frac{1}{2}}(\hat{\boldsymbol{\omega}}_A - \boldsymbol{\omega}_0) \xrightarrow{D}, N(\mathbf{0}, \mathbf{B}^{-1}\mathbf{G}\mathbf{B}^{-1}). \tag{3.4.5}$$

Proof. We have

$$\begin{aligned}
\frac{\partial l(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} &= \frac{\partial l_A(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = \mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}),\\
\frac{\partial l(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})}{\partial \sigma_\alpha^2} &= \frac{\partial l_A(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})}{\partial \sigma_\alpha^2} - \frac{1}{2}\operatorname{tr}\left[\left\{\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}\right\}^{-1}\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{Z}_1\mathbf{Z}_1^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}\right], \quad (3.4.6)\\
\frac{\partial l(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})}{\partial \sigma_e^2} &= \frac{\partial l_A(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})}{\partial \sigma_e^2} - \frac{1}{2}\operatorname{tr}\left[\left\{\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}\right\}^{-1}\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{Z}_2\mathbf{Z}_2^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}\right].
\end{aligned}$$

From lemma 18, we have

$$\mathbf{K}^{-\frac{1}{2}}\psi(\boldsymbol{\omega}) = \mathbf{K}^{-\frac{1}{2}}\psi_A(\boldsymbol{\omega}) + o(1). \quad (3.4.7)$$

Thus, we can write

$$\begin{aligned}
\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} \mathbf{K}^{-\frac{1}{2}}\psi_A(\boldsymbol{\omega}) &= \sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} \mathbf{K}^{-\frac{1}{2}}\psi(\boldsymbol{\omega}) + o(1)\\
&= \sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} \mathbf{K}^{-\frac{1}{2}}\left\{\psi(\boldsymbol{\omega}) - \psi(\boldsymbol{\omega}_0) - \mathrm{E}\left\{\psi(\boldsymbol{\omega}) - \psi(\boldsymbol{\omega}_0)\right\}\right\}\\
&\quad + \sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} \mathbf{K}^{-\frac{1}{2}}\left\{\psi(\boldsymbol{\omega}_0) - \xi(\boldsymbol{\theta}_0)\right\}\\
&\quad + \sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} \mathbf{K}^{-\frac{1}{2}}\left[\mathrm{E}\left\{\psi(\boldsymbol{\omega}) - \psi(\boldsymbol{\omega}_0) - \nabla\psi(\boldsymbol{\omega}_0)(\boldsymbol{\omega}-\boldsymbol{\omega}_0)\right\}\right]\\
&\quad + \sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} \mathbf{K}^{-\frac{1}{2}}\left[\xi(\boldsymbol{\theta}_0) + \mathrm{E}\left\{\nabla\psi(\boldsymbol{\omega}_0)\right\}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)\right] + o(1).
\end{aligned}$$

Applying (3.2.9), we obtain

$$\mathbf{K}^{-\frac{1}{2}}[\psi(\boldsymbol{\omega}_0) - \xi(\theta_0)] = o_p(1).$$

Also, combining lemma 16 and lemma 17, we obtain

$$\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} |\mathbf{K}^{-\frac{1}{2}}\left[\psi(\boldsymbol{\omega}) - \psi(\boldsymbol{\omega}_0) - \mathrm{E}\left\{\psi(\boldsymbol{\omega}) - \psi(\boldsymbol{\omega}_0)\right\}\right]| = o_p(1),$$

and

$$\sup_{|\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)|\le M} |\mathbf{K}^{-\frac{1}{2}}\left[\mathrm{E}\left\{\psi(\boldsymbol{\omega}) - \psi(\boldsymbol{\omega}_0)\right\} - \mathrm{E}\left\{\nabla\psi(\boldsymbol{\omega}_0)\right\}(\boldsymbol{\omega}-\boldsymbol{\omega}_0)\right]| = o_p(1).$$

Thus, on combining above equations, we can see that, uniformly on $\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0) \le M$,

$$\mathbf{K}^{-\frac{1}{2}}\psi_A(\boldsymbol{\omega}) = \mathbf{K}^{-\frac{1}{2}}\xi(\boldsymbol{\theta}_0) + \mathbf{B}_n\mathbf{K}^{\frac{1}{2}}(\boldsymbol{\omega}-\boldsymbol{\omega}_0) + o_p(1), \quad (3.4.8)$$

where $\mathbf{B}_n = -\mathbf{K}^{-\frac{1}{2}}\,\mathrm{E}\left\{\nabla\psi(\boldsymbol{\omega}_0)\right\}\mathbf{K}^{-\frac{1}{2}}$. The right-hand side of (3.4.8) is mainly determined by the negative of the quadratic form of $\mathbf{B}_n$ and is negative for M sufficiently large. Therefore, applying to Result 6.3.4 of Ortega and Rheinboldt (1973), we conclude that there exists a solution to the estimating equations exists in probability and satisfies $|\mathbf{K}^{-\frac{1}{2}}(\hat{\boldsymbol{\omega}}_A - \boldsymbol{\omega}_0)| = O_p(1)$. The asymptotic representation follows from lemma 16and lemma 17. Furthermore, the Lyapounov condition holds and hence the asymptotic normality result follows. □

3.5 Restricted Maximum Likelihood (REML) Estimating Equations

Turning to the residual log-likelihood function, we write

$$l_R(\mathbf{y};\boldsymbol{\theta}) = -\frac{1}{2}(n-r)\log 2\pi - \frac{1}{2}\log|\mathbf{V}(\boldsymbol{\theta})| - \frac{1}{2}\log|\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}| - \frac{1}{2}\mathbf{y}^T\mathbf{P}(\boldsymbol{\theta})\mathbf{y}, \quad (3.5.1)$$

where $\mathbf{P}(\boldsymbol{\theta}) = \mathbf{V}^{-1}(\boldsymbol{\theta}) - \mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}\{\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}\}^{-1}\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})$. We write the REML estimating equation as $\psi_R(\boldsymbol{\theta}) = 0$, where $\psi_R(\boldsymbol{\theta}) = [l_{R\sigma^2_\alpha}(\boldsymbol{\theta}), l_{R\sigma^2_e}(\boldsymbol{\theta})]^T$ with

$$\begin{aligned}
l_{R\sigma^2_\alpha}(\boldsymbol{\theta}) &= -\frac{1}{2}\operatorname{tr}\left\{\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{Z}_1\mathbf{Z}_1^T\right\} - \frac{1}{2}\operatorname{tr}\left[\{\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}\}^{-1}\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{Z}_1\mathbf{Z}_1^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}\right] \\
&\quad + \frac{1}{2}\mathbf{y}^T\mathbf{P}(\boldsymbol{\theta})\mathbf{Z}_1\mathbf{Z}_1^T\mathbf{P}(\boldsymbol{\theta})\mathbf{y}, \\
l_{R\sigma^2_e}(\boldsymbol{\theta}) &= -\frac{1}{2}\operatorname{tr}\left\{\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{Z}_2\mathbf{Z}_2^T\right\} - \frac{1}{2}\operatorname{tr}\left[\{\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}\}^{-1}\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{Z}_2\mathbf{Z}_2^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}\right] \\
&\quad + \frac{1}{2}\mathbf{y}^T\mathbf{P}(\boldsymbol{\theta})\mathbf{Z}_2\mathbf{Z}_2^T\mathbf{P}(\boldsymbol{\theta})\mathbf{y}.
\end{aligned} \quad (3.5.2)$$

We find that $l_R(\mathbf{y};\boldsymbol{\theta})$ is the profile version of $l_A(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})$. That is

$$l_R(\mathbf{y};\boldsymbol{\theta}) = \sup_{\boldsymbol{\beta}} \quad l_A(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta}) = l_A(\mathbf{y};\hat{\boldsymbol{\beta}},\boldsymbol{\theta}),$$

where $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \{\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}\}^{-1}\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{y}$ is obtained by setting the first equation of (3.4.6) to zero. The AMLE for $\boldsymbol{\theta}$, which is the second component of the pair $(\boldsymbol{\beta},\boldsymbol{\theta})$ that maximizes $l(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})$, is the maximizer of the profile likelihood function. Put simply then, the REML estimator is defined by "taking the supermum in two steps." Patefield (1977) showed that in parametric models, the inverse of the observed profile information is equal to the $\boldsymbol{\theta}$ component of the full observed inverse information.

To see this, we first obtain the derivative matrix from the AML estimating equation, $\mathbf{D}_A(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})$,

$$\mathbf{D}_A(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial^2 l_A(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T} & \frac{\partial^2 l_A(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\theta}^T} \\ \frac{\partial^2 l_A(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\beta}^T} & \frac{\partial^2 l_A(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T} \end{bmatrix}.$$

Edwards (1984) defines the information matrix $\mathbf{F}_A(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})$ as the inverse of the matrix of second partial derivatives of the log-likelihood function,

$$\mathbf{F}_A(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta}) = \mathbf{D}_A^{-1}(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{F}_{A\boldsymbol{\beta}\boldsymbol{\beta}}(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta}) & \mathbf{F}_{A\boldsymbol{\beta}\boldsymbol{\theta}}(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta}) \\ \mathbf{F}_{A\boldsymbol{\theta}\boldsymbol{\beta}}(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta}) & \mathbf{F}_{A\boldsymbol{\theta}\boldsymbol{\theta}}(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta}) \end{bmatrix}.$$

If this information matrix of the log likelihood function is divided into sub-matrices corresponding to the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, then it can easily be shown that the sub-information matrix related to $\boldsymbol{\theta}$ is given by

$$\mathbf{F}_{A\boldsymbol{\theta}\boldsymbol{\theta}}(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta}) = -\left[\frac{\partial^2 l_A(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T} - \frac{\partial^2 l_A(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\beta}^T}\left\{\frac{\partial^2 l_A(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right\}^{-1}\frac{\partial^2 l_A(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\theta}^T}\right]^{-1} \tag{3.5.3}$$

The maximized function is essentially the projection of the profile of the log-likelihood function onto the $\boldsymbol{\theta}$ axes. This is leads to the lemma from Patefield (1977) .

Lemma 19. *For each $\boldsymbol{\theta} \in \Theta$, the sub-information matrix $\mathbf{F}_{A\boldsymbol{\theta}\boldsymbol{\theta}}(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})$ with $\boldsymbol{\beta}$ replaced by $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \{\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}\}^{-1}\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{y}$ is equal to the information matrix of the profile (maximized) likelihood function.*

Proof. The proof follows directly from the definitions of the profile (maximized) likelihood function and $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$. For each $\boldsymbol{\theta} \in \Theta$, $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ is a solution of the fixed effects likelihood equations,

$$\left.\frac{\partial l_A(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})}{\partial\boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})} = 0. \tag{3.5.4}$$

Differentiating both sides of (3.5.4) with respect to $\boldsymbol{\theta}$, we obtain

$$\frac{\partial}{\partial\boldsymbol{\theta}}\left\{\left.\frac{\partial l_A(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})}{\partial\boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})}\right\} = 0.$$

To simplify the derivatives, according to the implicit function theorem, we have

$$\left.\frac{\partial^2 l_A(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\beta}^T}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})} + \frac{\partial\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\left\{\left.\frac{\partial l_A(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})}\right\} = 0,$$

from which we obtain

$$\frac{\partial\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = -\left\{\left.\frac{\partial^2 l_A(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\beta}^T}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})}\right\}\left\{\left.\frac{\partial l_A(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})}\right\}^{-1}. \tag{3.5.5}$$

When $\boldsymbol{\theta} \in \Theta$, we see that the profile (maximized) likelihood function is given by $l_R(\mathbf{y};\boldsymbol{\theta}) = l_A\{\mathbf{y};\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}),\boldsymbol{\theta}\}$. Differentiating the profile (maximized) likelihood function with respect to $\boldsymbol{\theta}$, we obtain

$$\frac{\partial l_R(\mathbf{y};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \left.\frac{\partial l_A(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})} + \frac{\partial\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\left\{\left.\frac{\partial l_A(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})}{\partial\boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})}\right\}. \tag{3.5.6}$$

Substituting (3.5.4) into (3.5.6), we get

$$\frac{\partial l_R(\mathbf{y};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \left.\frac{\partial l_A(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})}. \tag{3.5.7}$$

Differentiating both sides of (3.5.7) with respect to $\boldsymbol{\theta}$ again, we obtain the negative of the information matrix of the profile(maximized) likelihood function as

$$\frac{\partial^2 l_R(\mathbf{y};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T} = \left.\frac{\partial^2 l_A(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})} + \frac{\partial\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\left\{\left.\frac{\partial^2 l_A(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\theta}^T}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})}\right\}. \tag{3.5.8}$$

Substituting (3.5.5) into (3.5.8), we obtain

$$\begin{aligned}
&-\frac{\partial^2 l_R(\mathbf{y};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\\
&= -\Bigg[\left.\frac{\partial^2 l_A(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})} - \left\{\left.\frac{\partial^2 l_A(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\beta}^T}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})}\right\}\left\{\left.\frac{\partial l_A(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})}\right\}^{-1}\Bigg\{\left.\frac{\partial^2 l_A(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\theta}^T}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}}
\end{aligned}$$

We calculate the information matrix $\mathbf{F}_R(\mathbf{y};\boldsymbol{\theta})$ of $l_R(\mathbf{y};\boldsymbol{\theta})$,

$$\begin{aligned}
&\mathbf{F}_R(\mathbf{y};\boldsymbol{\theta})\\
&= -\left\{\frac{\partial^2 l_R(\mathbf{y};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right\}^{-1}\\
&= -\Bigg[\left.\frac{\partial^2 l_A(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})} - \left\{\left.\frac{\partial^2 l_A(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\beta}^T}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})}\right\}\left\{\left.\frac{\partial l_A(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})}\right\}^{-1}\Bigg\{\left.\frac{\partial^2 l_A(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\theta}^T}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}}\\
&= \mathbf{F}_{A\boldsymbol{\theta}\boldsymbol{\theta}}\{\mathbf{y};\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}),\boldsymbol{\theta}\}.
\end{aligned} \tag{3.5.9}$$

Thus, comparing (3.5.3) and (3.5.9), we can see that the sub-information matrix $\mathbf{F}_{A\boldsymbol{\theta}\boldsymbol{\theta}}$ with $\boldsymbol{\beta}$ replaced by $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ is equal to the information matrix of the profile (maximized) likelihood function. □

Note that we start by the angle of the between parameter $\boldsymbol{\omega}^{(b)}$ and the within parameter $\boldsymbol{\omega}^{(w)}$ before (3.4.8). Now, we change the angle into the fixed parameter $\boldsymbol{\theta}$ and the random parameter $\boldsymbol{\theta}$. From (3.4.7), we have

$$\mathbf{K}^{-\frac{1}{2}}\psi_A(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0) = \mathbf{K}^{-\frac{1}{2}}\xi(\boldsymbol{\theta}_0) + o_p(1).$$

Furthermore, from (3.5.7), we have

$$\psi_R(\boldsymbol{\theta}_0) = \psi_A\{\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0\}.$$

Combining the above results, we obtain

$$\mathbf{K}_R^{-\frac{1}{2}}\psi_R(\boldsymbol{\theta}_0) = \mathbf{K}_R^{-\frac{1}{2}}\xi_R(\boldsymbol{\theta}_0) + o_p(1),$$

where $\mathbf{K}_R = \mathrm{diag}(g^{\frac{1}{2}}, n^{\frac{1}{2}})$ and $\xi_R(\boldsymbol{\theta}_0) = [\xi_{\sigma_\alpha^2}(\boldsymbol{\theta}_0), \xi_{\sigma_e^2}(\boldsymbol{\theta}_0)]^T$ from (3.2.10). Moreover, we also have

$$\mathbf{G}_R = \lim_{g\to\infty}\lim_{m_L\to\infty} \mathrm{Var}\left\{\mathbf{K}_R^{-\frac{1}{2}}\xi_R(\boldsymbol{\theta}_0)\right\}.$$

Comparing this with (3.2.12), we see that $\mathbf{G}_R$ is the random parameter part of $\mathbf{G}$.

Thus, we can straightforwardly obtain the following main result on the asymptotic properties of the REML estimator.

Theorem 5. *Suppose Condition B holds. Then there exist a solution $\hat{\boldsymbol{\theta}}_R$ to the REML estimating equation $\psi_R(\boldsymbol{\theta}) = 0$, satisfying $|\mathbf{K}_R^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_R - \boldsymbol{\theta}_0)| = O_p(1)$, as $g \to \infty$ and $m_L \to \infty$. We have*

$$\mathbf{K}_R^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_R - \boldsymbol{\theta}_0) = \mathbf{B}_R^{-1}\mathbf{K}_R^{-\frac{1}{2}}\xi_R(\boldsymbol{\theta}_0) + o_p(1),$$

and

$$\mathbf{K}_R^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_R - \boldsymbol{\theta}_0) \xrightarrow{D}, N(\mathbf{0}, \mathbf{B}_R^{-1}\mathbf{G}_R\mathbf{B}_R^{-1}),$$

where $\mathbf{B}_R = \lim_{g,m_L\to\infty} \mathbf{K}_R^{-\frac{1}{2}}\,\mathrm{E}\left\{\nabla\psi_R(\boldsymbol{\theta}_0)\right\}\mathbf{K}_R^{-\frac{1}{2}}$.

Proof. We divide the AML estimator $\hat{\boldsymbol{\omega}}_A$ into the fixed parameters $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ and the random parameters $\hat{\boldsymbol{\theta}}_A$, and we also partition $\psi_A(\boldsymbol{\omega})$ into two components, $\psi_{AF}(\boldsymbol{\beta},\boldsymbol{\theta}) = [l_{A\beta_0}(\boldsymbol{\beta},\boldsymbol{\theta}), l_{A\beta_1}(\boldsymbol{\beta},\boldsymbol{\theta}), l_{A\beta_2}(\boldsymbol{\beta},\boldsymbol{\theta})]^T$ and $\psi_{AR}(\boldsymbol{\beta},\boldsymbol{\theta}) = [l_{A\sigma_\alpha^2}(\boldsymbol{\beta},\boldsymbol{\theta}), l_{A\sigma_e^2}(\boldsymbol{\beta},\boldsymbol{\theta})]^T$. It follows that $\nabla\psi_{AF}(\boldsymbol{\beta},\boldsymbol{\theta}) = \frac{\partial\psi_{AF}(\boldsymbol{\beta},\boldsymbol{\theta})}{\partial\boldsymbol{\beta}}$ and $\nabla\psi_{AR}(\boldsymbol{\beta},\boldsymbol{\theta}) = \frac{\partial\psi_{AR}(\boldsymbol{\beta},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}$.

For the fixed parameters, we let $\mathbf{B}_{AF} = \lim_{g\to\infty}\lim_{m_L\to\infty}\mathbf{K}_F^{-\frac{1}{2}}\,\mathrm{E}\left\{\nabla\psi_{AF}(\boldsymbol{\beta}_0,\boldsymbol{\theta}_0)\right\}\mathbf{K}_F^{-\frac{1}{2}}$ and $\mathbf{G}_{AF} = \lim_{g\to\infty}\lim_{m_L\to\infty}\mathrm{Var}\left\{\mathbf{K}_F^{-\frac{1}{2}}\xi_F(\boldsymbol{\theta}_0)\right\}$, where $\mathbf{K}_F = \mathrm{diag}(g^{\frac{1}{2}}, g^{\frac{1}{2}}, n^{\frac{1}{2}})$ and $\xi_F(\boldsymbol{\theta}_0) = [\xi_{\beta_0}(\boldsymbol{\theta}_0), \xi_{\beta_1}(\boldsymbol{\theta}_0), \xi_{\beta_2}(\boldsymbol{\theta}_0)]^T$ Thus, we can rewrite (3.4.4) and (3.4.5) as

$$\mathbf{K}_F^{\frac{1}{2}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \mathbf{B}_{AF}^{-1}\mathbf{K}_F^{-\frac{1}{2}}\xi_F(\boldsymbol{\theta}_0) + o_p(1) \tag{3.5.10}$$

and

$$\mathbf{K}_F^{\frac{1}{2}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{D}, N(\mathbf{0}, \mathbf{B}_{AF}^{-1}\mathbf{G}_{AF}\mathbf{B}_{AF}^{-1}). \tag{3.5.11}$$

For the random parameters, we denote $\mathbf{B}_{AR} = \lim_{g\to\infty} \lim_{m_L\to\infty} \mathbf{K}_R^{-\frac{1}{2}} \operatorname{E}\{\nabla\psi_{AR}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)\} \mathbf{K}_R^{-\frac{1}{2}}$ and $\mathbf{G}_{AR} = \lim_{g\to\infty} \lim_{m_L\to\infty} \operatorname{Var}\left\{\mathbf{K}_R^{-\frac{1}{2}} \xi_R(\boldsymbol{\theta}_0)\right\}$, where $\xi_R(\boldsymbol{\theta}_0) = [\xi_{\sigma_\alpha^2}(\boldsymbol{\theta}_0), \xi_{\sigma_e^2}(\boldsymbol{\theta}_0)]^T$. Thus, we can rewrite (3.4.4) and (3.4.5) as

$$\mathbf{K}_R^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_A - \boldsymbol{\theta}_0) = \mathbf{B}_{AR}^{-1}\mathbf{K}_R^{-\frac{1}{2}}\xi_R(\boldsymbol{\theta}_0) + o_p(1), \tag{3.5.12}$$

and

$$\mathbf{K}_R^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_A - \boldsymbol{\theta}_0) \xrightarrow{D}, N(\mathbf{0}, \mathbf{B}_{AR}^{-1}\mathbf{G}_{AR}\mathbf{B}_{AR}^{-1}). \tag{3.5.13}$$

From (3.5.7), we find that the REML estimating equation for $\boldsymbol{\theta}$ equals the AML estimating equation for $\boldsymbol{\theta}$ with $\boldsymbol{\beta}$ replaced by $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$. Therefore we obtain $\psi_R(\boldsymbol{\theta}) = \psi_{AR}\{\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}), \boldsymbol{\theta}\}$, and consequently, we can find $\hat{\boldsymbol{\theta}}_R = \hat{\boldsymbol{\theta}}_A$ satisfying

$$\psi_R(\hat{\boldsymbol{\theta}}_R) = \psi_{AR}\{\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}), \hat{\boldsymbol{\theta}}_A\} = 0.$$

We also have $\nabla\psi_R(\boldsymbol{\theta}) = \nabla\psi_{AR}\{\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}), \boldsymbol{\theta}\}$, so $\mathbf{B}_{AR} = \mathbf{B}_R$. The required result follows from this. $\square$

3.6 Discussion

We have established the asymptotic equivalence and asymptotic normality of the maximum likelihood and REML estimators of the parameters in the nested error regression model under very mild conditions when both the number of clusters and the cluster size increase to infinity but the rank of the design matrix for $\mathbf{X}$, is fixed (Jiang (1996)) .The results show explicitly that the between and within parameters are estimated at different rates and are asymptotically orthogonal. This means that within cluster covariates should have their between cluster contribution (as represented by the within cluster means) separated out and included in the model as contextual effects. One motivation for allowing the cluster size to increase with the number of clusters is that this is required for consistent prediction of the random effects. We have not considered prediction of the random effects explicitly in this chapter but will do so in next chapter. The model we have considered is a simple linear mixed model. It is of interest to extend our results to more general linear mixed models and indeed to generalized linear mixed models. We explained the conditions that are needed and described the form of the asymptotic variance of the estimators that we obtain when we have multiple covariates in the model. It is also of clear that we can extend the hierarchical structure of the model and allow for more variance components. The effect is to increase the sets of parameters (beyond just between and within) so that there is a set for each level in the hierarchy, with the parameters in each level converging at a different rate and the limit distribution having a diagonal block for each level in the hierarchy. Of course, this means that the covariates need to be appropriately centered at each level of the hierarchy and the centering means (themselves recenter) included as contextual effects in the higher level. We plan to treat this problem in subsequent work. Finally, the maximum likelihood and REML estimators are not the only estimators of interest for the parameters of linear mixed models. Other estimators (including robust estimators) are available and it is also of interest to derive their asymptotic properties. We expect that the form of the asymptotic covariance matrices for these estimators will be block diagonal with a separate block for the parameters at each level in the hierarchy, just as we found for the maximum likelihood and REML estimators.

Chapter 4 Asymptotic Properties of empirical Best Linear Unbiased Predictors (EBLUPs) for Hierarchical Linear Mixed Models with Large Unbalanced Cluster Sizes

Part of this chapter was submitted as a paper to a statistical journal for publication.

4.1 Introduction

Mixed models are a general class of statistical models that relate response variables to observed covariates or explanatory variables and to unobserved random terms called random effects. The random effects are included to capture sampling from a population, describe dependence, produce parsimonious models and/or improve prediction. The realized values of the random effects are explicitly important in some prediction problems, but are otherwise often treated as nuisance terms in the model. Since model fitting typically involves integrating (either exactly or approximately) the unobserved random effects over their distribution, even when they are not of direct interest, the random effects play an important, implicit role in fitting and using the model, so it is a good practice to obtain estimates or predictions of the random effects and use these to investigate their distributions.

In linear mixed models, the most widely used estimates of the realized values of the random effects are empirical best linear unbiased predictors (EBLUPs) which are estimated versions of best linear unbiased predictors (BLUPs) derived by Henderson et al. (1959); see for example Henderson (1975), Robinson (1991) and Jiang (1998). EBLUPs have been used in prediction of breeding values (Henderson, 1984) and small area means (Battese et al., 1988) as well as in many other application areas. Many predictions are presented as point predictions; others use the approximations to their prediction mean squared error under normality developed by Kackar and Harville (1984) and Prasad and Rao (1990) to give some measure of uncertainty. Diagnostic methods using EBLUPs are discussed in, for example, Lange and Ryan (1989), Verbeke and

Molenberghs (2000, ch 7), Pinheiro and Bates (2000, Sec 4.3.2), Nobre and Singer (2007), and Schützenmeister et al. (2012). In spite of their wide use in practice, very little is actually known about the theoretical properties of EBLUPs. Jiang (1998) gave conditions under which EBLUPs are consistent for the random effects and obtained a consistency result for the empirical distribution of the EBLUPs, but there are no results on the asymptotic distribution of EBLUPs. In this chapter we consider a particular linear mixed model and derive the asymptotic distributions of EBLUPs.

Now, we study the nested error regression model (3.1.1) from Chapter 3. We write

$$y_{ij} = \mu(x_{ij}) + \alpha_i + e_{ij}, \qquad j = 1, \ldots, m_i,\, i = 1, \ldots, g,$$

where $\mu(x_{ij})$ is the conditional mean of the response given the covariate x_{ij}, α_i is a random effect representing a random intercept or cluster effect and e_{ij} is an error term. We assume that the $\{\alpha_i\}$ and $\{e_{ij}\}$ are all mutually independent with mean zero and variances (called the variance components) σ_α^2 and σ_e^2, respectively. We do not assume that these random variables are normally distributed so the responses y_i are not necessarily normally distributed. Also, we use a dot above a parameter or function of the parameters to denote the true value of the parameter or function of the parameters.

In general, the mean function μ and the variance components σ_α^2 and σ_e^2 are unknown and are estimated from data. Estimating the parameters in the model under the asymptotic setting in which $m_L \to \infty$ and $g \to \infty$ is more complicated than doing so in the more familiar setting in which m_U is bounded as $g \to \infty$. From Chapter 3, we derived asymptotic results for maximum likelihood and restricted or residual maximum likelihood (REML) estimators of the parameters in the model (3.1.1) with a linear mean function as g and $m_L \to \infty$. We found that the mean function μ in the model (3.1.1) needs to be formulated in a particular way. Specifically, subtract the cluster means of the covariates that vary within clusters (within cluster covariates) from these covariates and then group the centered covariates into $x_{ij} - \bar{x}_{i.}$. The centering ensures that $\sum_{j=1}^{m_i}(x_{ij} - \bar{x}_{i.}) = 0$ for all $i = 1, \ldots, g$. Then group the means of the within cluster covariates that are not all identically zero (contextual effects) with all the remaining covariates (between cluster covariates that are constant within clusters) into a covariate that we denote $\bar{x}_{i.}$ and set

$$\mu(x_{ij}) = \begin{cases} \beta_0 + \beta_1 x_i & \text{if } x_{ij} = x_i \\ \beta_0 + \beta_1 \bar{x}_{i.} + \beta_2(x_{ij} - \bar{x}_{i.}) & \text{otherwise} \end{cases},$$

where β_0 is the unknown intercept, β_1 is the unknown between cluster slope parameter and β_2 is the unknown within cluster slope parameter. Centering the within cluster covariates is very important for the asymptotic properties of the maximum likelihood and REML estimators when

$m_i \to \infty$ because the maximum likelihood and REML estimators of within cluster slope parameters and between cluster slope parameters have different asymptotic behavior from Chapter 3.

Under the model (3.1.1) and (3.1.2), we can write the BLUPs α_i as

$$\tilde{\alpha}_i = \frac{m_i \dot{\sigma}_\alpha^2}{\dot{\sigma}_e^2 + m_i \dot{\sigma}_\alpha^2}(\bar{y}_{i.} - \beta_0 - \bar{x}_{i.}\beta_1). \tag{4.1.1}$$

The EBLUPs $\hat{\alpha}_i$ are the BLUPs $\tilde{\alpha}_i$ evaluated at the maximum likelihood or REML estimators of βs and $\boldsymbol{\theta} = (\sigma_\alpha^2, \sigma_e^2)^T$. We establish the asymptotic distribution of EBLUPs under the model (3.1.1) and (3.1.2) and very simple conditions which do not require normality. Our model has the simplest possible random effect structure but is nonetheless still useful in practice. The advantage of considering a simple model is that we can easily see how to derive the properties of EBLUPs and why their properties are what they are. Our results on the asymptotic distribution of EBLUPs lead to a very simple approximation to the mean squared error of EBLUPs that is easy to estimate and use to construct asymptotic prediction intervals for the random effects. Our approximation is so much simpler than those of Kackar and Harville (1984) and Prasad and Rao (1990) that we present both a detailed theoretical comparison and a simulation study to explain and justify the good performance of our approximation and its estimator.

Above of all, we study the special case of the model (3.1.1), the balanced one-way random model, in section 4.2. Next, we discuss maximum likelihood and REML estimation of the parameters of the model (3.1.1) in Section 4.3. We describe the central limit theorem for the maximum likelihood and REML estimators of the parameters as $g, m_L \to \infty$ established from Chapter 3 under simple conditions which do not require having to assume normality. Then in Section 4.3.1 we obtain central limit theorems for the EBLUPs and use these to develop approximations to the distribution of EBLUPs as the convolution of the distribution of the random effects and a normal distribution. We apply the asymptotic results from Section 4.3.1 in Section 4.3.2 to obtain very simple approximations to the mean squared error of EBLUPs. We discuss estimation of the mean squared errors and show how to construct asymptotic prediction intervals for the random effects. We present the results of a simulations study to compare the finite sample performance of our mean squared error estimator with those of Kackar and Harville (1984) and Prasad and Rao (1990) in Section 4.4. Finally, we conclude with a discussion in Section 4.5.

4.2 Balanced One-way Random Model

We consider the special case of model (3.1.1) in which the jth observation in the ith cluster has the representation

$$y_{ij} = \beta_0 + \alpha_i + e_{ij} \qquad i = 1, \dots, g; j = 1, \dots, m, \tag{4.2.1}$$

where β_0 is the intercept, α_i is a random effect with mean 0 and variance σ_α^2 and e_{ij} is an error term with mean 0 and variance σ_e^2. This model allows for tractable approximating expressions for the MSE of predictors of α_i.

4.2.1 Main results on REML estimators

To control the estimating function and derive the asymptotic properties of $\hat{\boldsymbol{\theta}}$ from the estimating equation, we impose the following condition.

Condition C

1. The model (4.2.1) holds with true parameters $\boldsymbol{\theta}$ inside the parameter space Θ.

2. The number of clusters $g \to \infty$ and the cluster size $m \to \infty$.

3. The random variables $\{\alpha_i\}$ and $\{e_{ij}\}$ are independent and identically distributed and there exists a $\delta > 0$ such that $\mathrm{E}\,|\alpha_i|^{4+\delta} < \infty$ and $\mathrm{E}\,|e_{ij}|^{4+\delta} < \infty$ for all $i = 1, \dots, g$ and $j = 1, \dots, m_i$.

These are mild conditions which are often satisfied in practice. Conditions C3 ensures that limits needed to ensure the existence of the asymptotic variance of the estimating function exist and that we can establish a Lyapunov condition and hence a central limit theorem for the estimating function.

Theorem 6. *Suppose Condition C holds. Then, as $g, m \to \infty$, there exists a REML estimator $\hat{\boldsymbol{\theta}}$, satisfying $|\mathbf{K}_R^{\frac{1}{2}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)| = O_p(1)$. Moreover, it follows that*

$$\begin{bmatrix} g^{-\frac{1}{2}}(\hat{\sigma}_\alpha^2 - \dot{\sigma_\alpha}^2) \\ g^{-\frac{1}{2}}m^{-\frac{1}{2}}(\hat{\sigma}_e^2 - \dot{\sigma_e}^2) \end{bmatrix} \xrightarrow{D} N\left(\mathbf{0}, \begin{bmatrix} \mathrm{E}\,\alpha_1^4 - \dot{\sigma_\alpha}^4 & 0 \\ 0 & \mathrm{E}\,e_{ij}^4 - \dot{\sigma_e}^4 \end{bmatrix}\right).$$

Proof. The proof is given in Chapter 3. □

We can estimate variance components, σ_α^2 and σ_e^2, consistently when $g \to \infty$ with m bounded. If $m \to \infty$ but g is held fixed, we can only estimate the within cluster variance σ_e^2 consistently but not the between cluster variance σ_α^2. The theorem motivates allowing both $g \to \infty$ and $m \to \infty$ to estimate both parameters of σ_α^2 and σ_e^2 consistently. In fact, Jiang (1997) showed that the consistency of the REML estimator, in the case of balanced data such as in this case, holds in a stronger sense, the so-call Wald consistency. See Theorem 3.1 of Jiang (1997).

4.2.2 Asymptotic properties of the EBLUP

Under the model, (4.2.1), the BLUP for α_i is given by Robinson (1991)

$$\tilde{\alpha}_i = \frac{m\sigma_\alpha^2}{(\sigma_e^2 + m\sigma_\alpha^2)}(\bar{y}_{i.} - \bar{y}_{..}). \tag{4.2.2}$$

Since the above expressions depend on the variance components $\boldsymbol{\theta} = (\sigma_\alpha^2, \sigma_e^2)^T$. When $\boldsymbol{\theta}$ is unknown, (4.2.2) is not calculable. It is customary to replace $\boldsymbol{\theta}$ by $\hat{\boldsymbol{\theta}}$, a consistent estimator, such as the ML or REML estimator by Harville (1977). The resulting expression is usually referred to as the estimated BLUP (EBLUP) from Harville (1991).

Now, we study the asymptotic distribution of the EBLUP, using the central limit theorem for the REML estimator of $\boldsymbol{\theta}$ under Condition C. Let $*$ denote the convolution of two distributions.

Theorem 7. *Suppose condition C holds. Then as $g, m \to \infty$, we have the following results:*

1. *when $m/g \to 0$, we have*
$$m^{\frac{1}{2}}(\hat{\alpha}_i - \alpha_i) \xrightarrow{D} N(0, \dot{\sigma_e}^2),$$
2. *when $g/m \to 0$, we have*
$$g^{\frac{1}{2}}(\hat{\alpha}_i - \alpha_i) \xrightarrow{D} N(0, \dot{\sigma_\alpha}^2),$$
3. *when $g/m \to r$, we have*
$$g^{\frac{1}{2}}(\hat{\alpha}_i - \alpha_i) \xrightarrow{D} N(0, r\dot{\sigma_e}^2 + \dot{\sigma_\alpha}^2),$$
where $0 < r < \infty$.

Remarks. The above results can be summarized succinctly by stating that for large m and g, we have the approximation
$$\hat{\alpha}_i - \alpha_i \sim N\Big(0, \frac{\dot{\sigma_e}^2}{m} + \frac{\dot{\sigma_\alpha}^2}{g}\Big).$$
Then since the leading term in the linear approximation to $\hat{\alpha}_i - \alpha_i$ can be written as

$$-\bar{\alpha} + \bar{e}_{i.} + O_p(g^{-\frac{1}{2}}m^{-\frac{1}{2}}).$$

It follows that approximately for large m and g,

$$\hat{\alpha}_i \sim F_\alpha * N\Big(0, \frac{\dot{\sigma_e}^2}{m} + \frac{\dot{\sigma_\alpha}^2}{g}\Big),$$

where $*$ denotes the convolution of two distributions. That is, the distribution of the EBLUP $\hat{\alpha}_i$ is approximately the convolution of the distribution F_α of α_i and a normal distribution. In the classical Gaussian version of the model, $F = N(0, \dot{\sigma}_\alpha^2)$ and the EBLUPs

$$\hat{\alpha}_i \sim N\Big(0, \dot{\sigma}_\alpha^2 + \frac{\dot{\sigma_e}^2}{m} + \frac{\dot{\sigma_\alpha}^2}{g}\Big).$$

The beauty of our result is that it holds very generally, even when the random effects are not normally distributed.

Proof. We calculate

$$\begin{aligned}
\hat{\alpha}_i &= \frac{m\hat{\sigma}_\alpha^2}{\hat{\sigma}_e^2 + m\hat{\sigma}_\alpha^2}(\bar{y}_{i.} - \bar{y}_{..}) \\
&= \frac{m\hat{\sigma}_\alpha^2}{\hat{\sigma}_e^2 + m\hat{\sigma}_\alpha^2}(\alpha_i - \bar{\alpha} + \bar{e}_{i.} - \bar{e}_{..}) \\
&= \frac{m\dot{\sigma}_\alpha^2}{\dot{\sigma}_e^2 + m\dot{\sigma}_\alpha^2}(\alpha_i - \bar{\alpha} + \bar{e}_{i.} - \bar{e}_{..}) + (\frac{m\hat{\sigma}_\alpha^2}{\hat{\sigma}_e^2 + m\hat{\sigma}_\alpha^2} - \frac{m\dot{\sigma}_\alpha^2}{\dot{\sigma}_e^2 + m\dot{\sigma}_\alpha^2})(\alpha_i - \bar{\alpha} + \bar{e}_{i.} - \bar{e}_{..}) \quad (4.2.3) \\
&= \alpha_i - \bar{\alpha} + \bar{e}_{i.} - \bar{e}_{..} - \frac{\dot{\sigma_e}^2}{\dot{\sigma_e}^2 + m\dot{\sigma_\alpha}^2}(\alpha_i - \bar{\alpha} + \bar{e}_{i.} - \bar{e}_{..}) \\
&\quad + \frac{m\dot{\sigma_e}^2(\hat{\sigma}_\alpha^2 - \dot{\sigma_\alpha}^2) + m\dot{\sigma_\alpha}^2(\hat{\sigma}_e^2 - \dot{\sigma_e}^2)}{(\dot{\sigma_e}^2 + m\dot{\sigma_\alpha}^2)(\hat{\sigma}_e^2 + m\hat{\sigma}_\alpha^2)}(\alpha_i - \bar{\alpha} + \bar{e}_{i.} - \bar{e}_{..}).
\end{aligned}$$

Using the asymptotic properties of REML estimator, $\hat{\boldsymbol{\theta}}$ from Theorem 6, when $g, m \to \infty$,we have

$$\begin{aligned}
&\bar{\alpha} = O_p(g^{-\frac{1}{2}}), \quad \bar{e}_{i.} = O_p(m^{-\frac{1}{2}}), \quad , \bar{e}_{..} = O_p(g^{-\frac{1}{2}}m^{-\frac{1}{2}}), \\
&\hat{\sigma}_\alpha^2 - \dot{\sigma_\alpha}^2 = O_p(g^{-\frac{1}{2}}), \quad \text{and} \quad \hat{\sigma}_e^2 - \dot{\sigma_e}^2 = O_p(g^{-\frac{1}{2}}m^{-\frac{1}{2}}).
\end{aligned}$$

Then, we have an approximation for $\hat{\alpha}_i$

$$\hat{\alpha}_i = \alpha_i - \bar{\alpha} + \bar{e}_{i.} + O_p(g^{-\frac{1}{2}}m^{-\frac{1}{2}}). \quad (4.2.4)$$

When $g \gg m$, we have

$$\hat{\alpha}_i = \alpha_i + \bar{e}_{i.} + O_p(g^{-\frac{1}{2}}).$$

When $g \ll m$, we have

$$\hat{\alpha}_i = \alpha_i - \bar{\alpha} + O_p(m^{-\frac{1}{2}}) = (1 - \frac{1}{g})\alpha_i - \frac{1}{g}\sum_{k\neq i}\alpha_k + O_p(m^{-\frac{1}{2}}).$$

Moreover, let $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_g)^T$, and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_g)^T$. From the approximation (4.2.4), we have

$$\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} = -\begin{pmatrix} \bar{\alpha} - \bar{e}_{1.} \\ \vdots \\ \bar{\alpha} - \bar{e}_{g.} \end{pmatrix} + o_p(1).$$

Thus, suppose Condition C holds, we obtain

$$\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} \xrightarrow{D} N(\mathbf{0}, \frac{\dot{\sigma_\alpha}^2}{g}\mathbf{J}_g + \frac{\dot{\sigma_e}^2}{m}\mathbf{I}_g).$$

□

4.2.3 Mean squared error of the EBLUP

For the mean squared error(MSE) of the EBLUP, from Theorem 7 and (4.2.4), we have our own approximation

$$\mathrm{MSE}_{LW} = \mathrm{E}(\hat{\alpha}_i - \alpha_i)^2 = \frac{\dot{\sigma_\alpha}^2}{g} + \frac{\dot{\sigma_e}^2}{m} + o(1). \tag{4.2.5}$$

Kackar and Harville (1984) proposed an approximation for the MSE of the EBLUP under normal distribution. It also requires some stronger conditions including its third-order central moments equal zero. In (4.2.1), Kackar and Harville (1984) gave an approximation to the MSE,

$$\begin{aligned}\mathrm{MSE}_{KH} &= \mathrm{E}(\tilde{\alpha}_i - \alpha_i)^2 + \mathrm{E}(\hat{\alpha}_i - \tilde{\alpha}_i)^2 \\ &= \frac{\dot{\sigma_\alpha}^2}{g} + \frac{\dot{\sigma_e}^2}{m} - \frac{\dot{\sigma_e}^2}{gm}\frac{m\dot{\sigma_\alpha}^2}{(\dot{\sigma_e}^2 + m\dot{\sigma_\alpha}^2)} - \frac{\dot{\sigma_e}^2}{m\dot{\sigma_e}^2 + m^2\dot{\sigma_\alpha}^2} + \frac{2(gm-1)}{g^2m(m-1)}\frac{\dot{\sigma_e}^4}{(\dot{\sigma_e}^2 + m\dot{\sigma_\alpha}^2)} + o(1).\end{aligned} \tag{4.2.6}$$

Prasad and Rao (1990) proposed another approximation for the MSE of the EBLUP based on Kackar and Harville (1984). Prasad and Rao (1990) changed conditions but still required normality and the approximation to $\mathrm{E}(\hat{\alpha}_i - \tilde{\alpha}_i)^2$.

$$\begin{aligned}\mathrm{MSE}_{PR} &= \mathrm{E}(\tilde{\alpha}_i - \alpha_i)^2 + \mathrm{E}(\hat{\alpha}_i - \tilde{\alpha}_i)^2 \\ &= \frac{\dot{\sigma_\alpha}^2}{g} + \frac{\dot{\sigma_e}^2}{m} - \frac{\dot{\sigma_e}^2}{gm}\frac{m\dot{\sigma_\alpha}^2}{(\dot{\sigma_e}^2 + m\dot{\sigma_\alpha}^2)} - \frac{\dot{\sigma_e}^2}{m\dot{\sigma_e}^2 + m^2\dot{\sigma_\alpha}^2} + \frac{2}{g(m-1)}\frac{\dot{\sigma_e}^4}{(\dot{\sigma_e}^2 + m\dot{\sigma_\alpha}^2)} + o(1).\end{aligned} \tag{4.2.7}$$

We notice that MSE_{LW} is able to be applied to more general case (non-normal case), but both the MSE_{KH} and MSE_{PR} are only applied to normal case. We calculate

$$\begin{aligned}\mathrm{MSE}_{LW} - \mathrm{MSE}_{KH} &= \frac{\dot{\sigma_e}^2}{(\dot{\sigma_e}^2 + m\dot{\sigma_\alpha}^2)}\{\frac{\dot{\sigma_\alpha}^2}{g} + \frac{1}{m} - \frac{2(gm-1)\dot{\sigma_e}^2}{g^2m(m-1)}\} \\ &= O(\frac{1}{gm}) + O(\frac{1}{m^2}) + O(\frac{1}{gm^2})\end{aligned} \tag{4.2.8}$$

and

$$\begin{aligned}\mathrm{MSE}_{LW} - \mathrm{MSE}_{PR} &= \frac{m\dot{\sigma_\alpha}^2}{(\dot{\sigma_e}^2 + m\dot{\sigma_\alpha}^2)}\{\frac{\dot{\sigma_\alpha}^2}{g} + \frac{1}{m} - \frac{2\dot{\sigma_e}^2}{g(m-1)}\} \\ &= O(\frac{1}{gm}) + O(\frac{1}{m^2}) + O(\frac{1}{gm^2}).\end{aligned} \tag{4.2.9}$$

We can find that our approximation has small difference with Kackar and Harville (1981) and Prasad and Rao (1990), as $g, m \to \infty$. When $g \gg m$, we have $\mathrm{MSE}_{LW} - \mathrm{MSE}_{KH} = O(\frac{1}{m^2})$ and $\mathrm{MSE}_{LW} - \mathrm{MSE}_{PR} = O(\frac{1}{m^2})$. When $g \ll m$, we have $\mathrm{MSE}_{LW} - \mathrm{MSE}_{KH} = o(1)$ and $\mathrm{MSE}_{LW} - \mathrm{MSE}_{PR} = o(1)$.

When the last term of (4.2.8) ,$\frac{\dot{\sigma_\alpha}^2}{g} + \frac{1}{m} - \frac{2(gm-1)\dot{\sigma_e}^2}{g^2m(m-1)} > 0$, we calculate

$$g\frac{m-1}{m} > 2(1 - \frac{1}{gm})\dot{\sigma_e}^2 - (m-1)\dot{\sigma_\alpha}^2. \tag{4.2.10}$$

We can find that for large m, our approximation is larger than the approximation of Kackar and Harville (1984). Similarly, when the last term of (4.2.9) ,$\frac{\dot{\sigma_\alpha}^2}{g} + \frac{1}{m} - \frac{2\dot{\sigma_e}^2}{g(m-1)} > 0$, we calculate

$$g\frac{m-1}{m} > 2\dot{\sigma_e}^2 - (m-1)\dot{\sigma_\alpha}^2. \tag{4.2.11}$$

We can find that for large m, our approximation is larger than the approximation of Prasad and Rao (1990).

When we get a large sample size (a large number of independent clusters and large cluster sizes), the difference between our approximation and MSE_{KH} and MSE_{PR} is minimal. However, in our approximation, it is not necessary to suppose that the model under the normality and the Condition A we set are much milder and often satisfied in practice. Balanced one-way random model is a straightforward case of the nested error regression model so that we can easily calculate the approximation to the MSE of EBLUP from Kackar and Harville (1984) and Prasad and Rao (1990) then check the difference with our approximation. In general case (non-normality), when we calculate approximations to the MSE from Kackar and Harville (1984) and Prasad and Rao (1990), it always requires us to check whether if it satisfies with their conditions and also needs to ensure that the model is under the normality. However, from (4.2.8) and (4.2.9), we still can use approximations of Kackar and Harville (1984) and Prasad and Rao (1990) without their required conditions (including normality) as $g, m \to \infty$.

4.3 Nested error regression model

In this section, we consider the general nested error regression model (3.1.1) with the conditional mean $\mu(x_{ij})$ from (3.1.2)

$$\mu(x_{ij}) = \begin{cases} \beta_0 + \beta_1 x_i & \text{if } x_{ij} = x_i \\ \beta_0 + \beta_1 \bar{x}_{i.} + \beta_2(x_{ij} - \bar{x}_{i.}) & \text{otherwise} \end{cases} \quad \text{for} \quad i = 1, \dots, g \quad \text{and} j = 1, \dots, m_i,$$

where the βs are unknown coefficients, called regression parameters or fixed effects, and $\bar{x}_{i.} = m_i^{-1} \sum_{j=1}^{m_i} x_{ij}$ is the within cluster mean of the explanatory variable.

Under the model, the BLUP for α_i is given by

$$\tilde{\alpha}_i = \frac{m_i \dot{\sigma}_\alpha^2}{\dot{\sigma}_e^2 + m_i \dot{\sigma}_\alpha^2} (\bar{y}_{i.} - \beta_0 - \bar{x}_{i.} \beta_1). \tag{4.3.1}$$

When βs, σ_α^2 and σ_e^2 are unknown, (4.3.1) is not computable. It is customary to replace them by consistent estimates such as ML/ REML . The resulting expression is usually referred to as estimated BLUP (EBLUP) from Harville (1991). We get the EBLUP,

$$\hat{\alpha}_i = \frac{m_i \hat{\sigma}_\alpha^2}{\hat{\sigma}_e^2 + m_i \hat{\sigma}_\alpha^2} (\bar{y}_{i.} - \hat{\beta}_0 - \bar{x}_{i.} \hat{\beta}_1) \tag{4.3.2}$$

Note that the EBLUP is no longer a linear function of the observations y_{ij}. We first prove a central limit theorem for the variance components under mild conditions.

Theorem 8. *Suppose Condition B holds. Then Maximum likelihood and REML estimators* $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)^T$ *and* $\hat{\boldsymbol{\theta}} = (\hat{\sigma}_\alpha^2, \hat{\sigma}_e^2)^T$*, satisfy* $|\mathbf{K}_{f_1}^{\frac{1}{2}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)| = O_p(1)$ *and* $|\mathbf{K}_{f_2}^{\frac{1}{2}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)| = O_p(1)$*, respectively, as* $g, m_L \to \infty$*, where* $\mathbf{K}_{f_1} = \mathrm{diag}[g, g, n]$ *and* $\mathbf{K}_{f_2} = \mathrm{diag}[g, n]$*. Moreover,*

$$\mathbf{K}_{f_1}^{\frac{1}{2}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{D} N(\mathbf{0}, \mathbf{C}_{f1}), \quad \textit{and} \quad \mathbf{K}_{f_2}^{\frac{1}{2}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{D} N(\mathbf{0}, \mathbf{C}_{f2}),$$

where

$$\mathbf{C}_{f_1} = \begin{bmatrix} \frac{c_2 \dot{\sigma_\alpha}^2}{(c_2 - c_1^2)} & \frac{-c_1 \dot{\sigma_\alpha}^2}{(c_2 - c_1^2)} & 0 \\ \frac{-c_1 \dot{\sigma_\alpha}^2}{(c_2 - c_1^2)} & \frac{\dot{\sigma_\alpha}^2}{(c_2 - c_1^2)} & 0 \\ 0 & 0 & \frac{\dot{\sigma_e}^2}{c_3} \end{bmatrix} \quad \textit{and} \quad \mathbf{C}_{f_2} = \begin{bmatrix} \mathrm{E}\, \alpha_1^4 - \dot{\sigma_\alpha}^4 & 0 \\ 0 & \mathrm{E}\, e_{ij}^4 - \dot{\sigma_e}^4 \end{bmatrix}.$$

The proof is given in Chapter 3. From theorem 8, we can say that

$$\begin{aligned}&\hat{\beta}_0 - \dot{\beta}_0 = O_p(g^{-\frac{1}{2}}), \quad \hat{\beta}_1 - \dot{\beta}_1 = O_p(g^{-\frac{1}{2}}), \quad \hat{\beta}_2 - \dot{\beta}_2 = O_p(n^{-\frac{1}{2}}),\\ &\hat{\sigma}_\alpha - \dot{\sigma_\alpha} = O_p(g^{-\frac{1}{2}}), \quad \text{and} \quad \hat{\sigma}_e - \dot{\sigma_e} = O_p(n^{-\frac{1}{2}}).\end{aligned} \tag{4.3.3}$$

4.3.1 Asymptotic property of the EBLUP

Now, we study the asymptotic distribution of the EBLUP, using the central limit theorem for maximum likelihood and REML estimators under Condition B.

Theorem 9. *Suppose condition B holds. Then as $g, m_L \to \infty$, we have the following results:*

1. *when $m_i/g \to 0$, we have*
$$m_i^{\frac{1}{2}}(\hat{\alpha}_i - \alpha_i) \xrightarrow{D} N(0, \dot{\sigma_e}^2),$$

2. *when $g/m_i \to 0$, we have*
$$g^{\frac{1}{2}}(\hat{\alpha}_i - \alpha_i) \xrightarrow{D} N(0, H_i),$$
where $H_i = \frac{\dot{\sigma_\alpha}^2}{c_2 - c_1^2}(\bar{x}_{i.}^2 + c_2 - 2\bar{x}_{i.}c_1)$;

3. *when $g/m_i \to r_i$, we have*
$$g^{\frac{1}{2}}(\hat{\alpha}_i - \alpha_i) \xrightarrow{D} N(0, r_i\sigma_e^2 + H_i),$$
where $0 < r_i < \infty$.

Remarks. The above results can be summarized succinctly by stating that for large m_L and g, we have the approximation
$$\hat{\alpha}_i - \alpha_i \sim N\Big(0, \frac{\dot{\sigma_e}^2}{m_i} + \frac{H_i}{g}\Big).$$
Then since the leading term in the linear approximation to $\hat{\alpha}_i - \alpha_i$ can be written as

$$\bar{e}_{i.} + (\dot{\beta}_0 - \hat{\beta}_0) + \bar{x}_{i.}(\dot{\beta}_1 - \hat{\beta}_1) + O_p(g^{-\frac{1}{2}}m_i^{-1}).$$

It follows that approximately for large m_L and g,

$$\hat{\alpha}_i \sim F_\alpha * N\Big(0, \frac{\dot{\sigma_e}^2}{m_i} + \frac{H_i}{g}\Big),$$

where $*$ denotes the convolution of two distributions. That is, the distribution of the EBLUP $\hat{\alpha}_i$ is approximately the convolution of the distribution F_α of α_i and a normal distribution. In the classical Gaussian version of the model, $F = N(0, \dot{\sigma}_\alpha^2)$ and the EBLUPs

$$\hat{\alpha}_i \sim N\Big(0,\, \dot{\sigma}_\alpha^2 + \frac{\dot{\sigma_e}^2}{m_i} + \frac{H_i}{g}\Big).$$

The beauty of our result is that it holds very generally, even when the random effects are not normally distributed.

Proof. From (4.3.2), we calculate

$$\begin{aligned}
\hat{\alpha}_i &= \frac{m_i\hat{\sigma}_\alpha^2}{\hat{\sigma}_e^2 + m_i\hat{\sigma}_\alpha^2}\left\{\alpha_i + \bar{e}_{i.} + (\dot{\beta}_0 - \hat{\beta}_0) + \bar{x}_{i.}(\dot{\beta}_1 - \hat{\beta}_1)\right\} \\
&= \frac{m_i\dot{\sigma_\alpha}^2}{\dot{\sigma_e}^2 + m_i\dot{\sigma_\alpha}^2}\left\{\alpha_i + \bar{e}_{i.} + (\dot{\beta}_0 - \hat{\beta}_0) + \bar{x}_{i.}(\dot{\beta}_1 - \hat{\beta}_1)\right\} \\
&\quad + \left(\frac{m_i\hat{\sigma}_\alpha^2}{\hat{\sigma}_e^2 + m_i\hat{\sigma}_\alpha^2} - \frac{m_i\dot{\sigma_\alpha}^2}{\dot{\sigma_e}^2 + m_i\dot{\sigma_\alpha}^2}\right)\left\{\alpha_i + \bar{e}_{i.} + (\dot{\beta}_0 - \hat{\beta}_0) + \bar{x}_{i.}(\dot{\beta}_1 - \hat{\beta}_1)\right\}.
\end{aligned}$$

Using the asymptotic properties of ML and REML estimators, (4.3.3), as $g, m_L \to \infty$, we have an approximation for $\hat{\alpha}_i$

$$\hat{\alpha}_i = \alpha_i + \bar{e}_{i.} + (\dot{\beta}_0 - \hat{\beta}_0) + \bar{x}_{i.}(\dot{\beta}_1 - \hat{\beta}_1) + O_p(g^{-\frac{1}{2}}m_i^{-1}). \tag{4.3.4}$$

When $g \gg m_U$, we have the approximation

$$\hat{\alpha}_i = \alpha_i + \bar{e}_{i.} + O_p(g^{-\frac{1}{2}}).$$

When $g \ll m_L$, we have the approximation

$$\hat{\alpha}_i = \alpha_i + (\dot{\beta}_0 - \hat{\beta}_0) + \bar{x}_{i.}(\dot{\beta}_1 - \hat{\beta}_1) + O_p(m_i^{-\frac{1}{2}}).$$

Moreover, let $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_g)^T$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_g)^T$. From the approximation (4.3.4), we have

$$\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} = (\dot{\beta}_0 - \hat{\beta}_0)\mathbf{1}_g + (\dot{\beta}_1 - \hat{\beta}_1)\bar{\mathbf{x}} + \bar{\mathbf{e}} + O_p(g^{-\frac{1}{2}}m_L^{-1}),$$

where $\bar{\mathbf{x}} = (\bar{x}_1, \ldots, \bar{x}_g)^T$ and $\bar{\mathbf{e}} = (\bar{e}_{1.}, \ldots, \bar{e}_{g.})^T$. Thus, suppose Condition B holds, we obtain

$$\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} \xrightarrow{D} N\left(\mathbf{0}, \frac{\dot{\sigma_\alpha}^2}{g(c_2 - c_1^2)}(c_2\mathbf{J}_g + \bar{\mathbf{x}}\bar{\mathbf{x}}^T - 2c_1\mathbf{1}_g\bar{\mathbf{x}}^T) + \frac{\dot{\sigma_e}^2}{m_L}\mathbf{I}_g\right).$$

□

4.3.2 Mean squared error of the EBLUP

For the mean squared error (MSE) of the EBLUP, from Theorem 9 and (4.3.4), we obtain our own approximation

$$\mathrm{MSE}_{LW} = \mathrm{E}(\hat{\alpha}_i - \alpha_i)^2 = \frac{H_i}{g} + \frac{\dot{\sigma_e}^2}{m_i} \tag{4.3.5}$$

where $\hat{H}_i = \frac{\hat{\sigma}_\alpha^2}{c_2 - c_1^2}(c_2 + \bar{x}_{i.}^2 - 2\bar{x}_{i.}c_1)$. We estimated the mean squared error of the EBLUP as

$$\widehat{\mathrm{MSE}}_{LW} = \frac{\hat{H}_i}{g} + \frac{\hat{\sigma}_e^2}{m_i}$$

Theorem 10. *Suppose condition B holds. Then as $g, m_L \to \infty$, we have*

$$g^{\frac{1}{2}}(\widehat{\mathrm{MSE}}_{LW} - \mathrm{MSE}_{LW}) \xrightarrow{P} 0. \tag{4.3.6}$$

Proof. Since $n = \sum_{i=1}^{g} m_i \geq g m_L$, we have $n^{-\frac{1}{2}} \leq g^{-\frac{1}{2}} m_L^{-\frac{1}{2}}$. Then we obtain

$$\begin{aligned}
g^{\frac{1}{2}}(\widehat{\mathrm{MSE}}_{LW} - \mathrm{MSE}_{LW}) &= \frac{\hat{H}_i - H_i}{g^{\frac{1}{2}}} + \frac{g^{\frac{1}{2}}(\hat{\sigma}_e^2 - \dot{\sigma_e}^2)}{m_i} + O(g^{-\frac{1}{2}} m_i^{-2}) \\
&= \frac{(c_2 + \bar{x}_{i.}^2 - 2\bar{x}_{i.}c_1)}{c_2 - c_1^2} \frac{(\hat{\sigma}_\alpha^2 - \dot{\sigma_\alpha}^2)}{g^{\frac{1}{2}}} + \frac{g^{\frac{1}{2}}(\hat{\sigma}_e^2 - \dot{\sigma_e}^2)}{m_i} + O(g^{-\frac{1}{2}} m_i^{-2}) \\
&\leq O_p(g^{-1}) + O_p(m_L^{-\frac{3}{2}}) + O(g^{-\frac{1}{2}} m_i^{-2}).
\end{aligned}$$

□

The model (3.1.1) is a general nested error regression model involving covariate x_{ij}, which makes computing approximations of Kackar and Harville (1984) and Prasad and Rao (1990) more complicated. So we can not obtain the final result explicitly. We use some limited simulations to explore the performance of our approximations compared with Kackar and Harville (1984) and Prasad and Rao (1990) in the context of the nested error regression model (3.1.1).

4.4 Simulations

4.4.1 Simulation Design

For the general nested error regression model (3.1.1), it is complicated to obtain exact expressions of the approximations from Kackar and Harville (1984) and Prasad and Rao (1990). We denote $\mathbf{y}^T = (\mathbf{y}_1^T, \ldots, \mathbf{y}_g^T)$ and $\mathbf{X} = [\mathbf{1}_n, \bar{\mathbf{x}}, \mathbf{x} - \bar{\mathbf{x}}]$, where $\bar{\mathbf{x}} = [\mathbf{1}_{m_1}^T \bar{x}_{1.}, \ldots, \mathbf{1}_{m_g} \bar{x}_{g.}]$ and $\mathbf{x}^T = [\mathbf{x}_1^T, \ldots, \mathbf{x}_g^T]$. We generate data from the model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \mathbf{e}, \tag{4.4.1}$$

where $\mathbf{Z} = \text{diag}\{\mathbf{1}_{m_1}, \ldots, \mathbf{1}_{m_g}\}$ and $\mathbf{e} = (e_{11}, \ldots, e_{gm_g})^T$. Denoting $\boldsymbol{\theta} = (\sigma_\alpha^2, \sigma_e^2)^T$, we have

$$\mathbf{V}(\boldsymbol{\theta}) = \mathbf{Z}\mathbf{D}(\boldsymbol{\theta})\mathbf{Z}^T + \mathbf{R}(\boldsymbol{\theta}), \quad \mathbf{D}(\boldsymbol{\theta}) = \sigma_\alpha^2 \mathbf{I}_g, \quad \text{and} \quad \mathbf{R}(\boldsymbol{\theta}) = \sigma_e^2 \mathbf{I}_n.$$

The model (4.4.1) is a matrix form of the model (3.1.1) to benefit of exact expressions of the approximations from Kackar and Harville (1984) and Prasad and Rao (1990). In simulation settings, We generate x_{ij} as the rounded (to the nearest integer) parts of absolute normal $|N(0,16)|$ random variables. We set $\boldsymbol{\beta} = (5,7,3)^T$. We also take $\mathbf{e}$ to be normal distribution, $e_{ij} \sim N(0,49)$ and $e_{ij} \sim N(0,96)$. We consider the distribution of the random effect $\boldsymbol{\alpha}$ to be normal distribution and mixture normal distribution, separately. For the normal distribution of the random effect $\boldsymbol{\alpha}$, we set $\alpha_i \sim N(0,25)$ and $\alpha_i \sim N(0,81)$. For the mixture of normal distribution of the random effect, let $\boldsymbol{\alpha}$ be a two-component mixture model,

$$\alpha_i \sim kN(\mu_{1\alpha}, \sigma_{1\alpha}^2) + (1-k)N(\mu_{2\alpha}, \sigma_{2\alpha}^2), \tag{4.4.2}$$

where k is the mixing weights. we have

$$\mathrm{E}\,\alpha_i = k\mu_{1\alpha} + (1-k)\mu_{2\alpha} \quad \text{and} \quad \mathrm{Var}(\alpha_i) = k(\mu_{1\alpha}^2 + \sigma_{1\alpha}^2) + (1-k)(\mu_{2\alpha}^2 + \sigma_{2\alpha}^2) - \mathrm{E}^2\,\alpha_i.$$

In the simulation, let mixing weight $k = 0.3$ and $\mu_{1\alpha} = 14$ and $\mu_{2\alpha} = -6$, which ensures $\mathrm{E}\,\alpha_i = 0$. And we set two pairs of variance components, $(\sigma_{1\alpha}^2 = 25, \sigma_{2\alpha}^2 = 81)$ and $(\sigma_{1\alpha}^2 = 9, \sigma_{2\alpha}^2 = 36)$. These choices ensure that $\sigma_\alpha^2 = \mathrm{Var}(\alpha_i) = 148.2$ or 111.9. The following histograms are data generated from two different mixture distributions simulation settings.

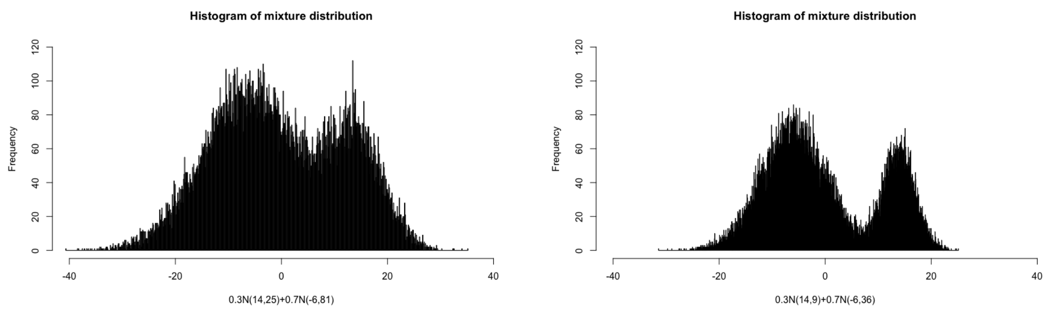

Figure 4.1: Histograms of mixture distribution

In this model, we obtain the BLUP of α_i,

$$\tilde{\alpha}_i(\boldsymbol{\theta}) = \boldsymbol{\lambda}_i^T \mathbf{D}(\boldsymbol{\theta})\mathbf{Z}^T \mathbf{P}(\boldsymbol{\theta})\mathbf{y},$$

where $\boldsymbol{\lambda}_i$ is a $g \times 1$ vector of zeros with a one in the i-th position and $\mathbf{P}(\boldsymbol{\theta}) = \mathbf{V}^{-1}(\boldsymbol{\theta}) -$

$\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}\{\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}\}^{-1}\mathbf{X}^T\mathbf{V}^{-1}(\boldsymbol{\theta})$. The EBLUPs, $\hat{\alpha}_i$ are obtained by

$$\hat{\alpha}_i(\boldsymbol{\theta}) = \boldsymbol{\lambda}_i^T\mathbf{D}(\hat{\boldsymbol{\theta}})\mathbf{Z}^T\mathbf{P}(\hat{\boldsymbol{\theta}})\mathbf{y},$$

Our approximation to the MSE of $\hat{\alpha}_i(\boldsymbol{\theta})$ is

$$\text{MSE}_{LW} = \text{E}\{\hat{\alpha}_i(\boldsymbol{\theta}) - \alpha_i\}^2 = \frac{H_i}{g} + \frac{\dot{\sigma_e}^2}{m_i}.$$

Kackar and Harville (1984) proposed an approximation to the MSE of the EBLUP as follow

$$\text{MSE}_{KH} = \text{E}\{\tilde{\boldsymbol{\alpha}}_i(\boldsymbol{\theta}) - \boldsymbol{\alpha}_i\}^2 + \text{E}\{\hat{\boldsymbol{\alpha}}_i(\boldsymbol{\theta}) - \tilde{\boldsymbol{\alpha}}_i(\boldsymbol{\theta})\}^2 = M_1(\boldsymbol{\theta}) + M_{KH_2}(\boldsymbol{\theta}), \tag{4.4.3}$$

where

$$\begin{aligned} M_1(\boldsymbol{\theta}) &= \text{Var}\{\tilde{\boldsymbol{\alpha}}_i(\boldsymbol{\theta}) - \boldsymbol{\alpha}_i\} \\ &= \boldsymbol{\lambda}_i^T\mathbf{D}(\boldsymbol{\theta})\mathbf{Z}^T\mathbf{P}(\boldsymbol{\theta})\mathbf{V}(\boldsymbol{\theta})\mathbf{P}(\boldsymbol{\theta})\mathbf{Z}\mathbf{D}(\boldsymbol{\theta})\boldsymbol{\lambda}_i + \boldsymbol{\lambda}_i^T\mathbf{D}(\boldsymbol{\theta})\boldsymbol{\lambda}_i - 2\boldsymbol{\lambda}_i^T\mathbf{D}(\boldsymbol{\theta})\mathbf{Z}^T\mathbf{P}(\boldsymbol{\theta})\mathbf{Z}\mathbf{D}(\boldsymbol{\theta})\boldsymbol{\lambda}_i, \end{aligned}$$

and

$$M_{KH_2}(\boldsymbol{\theta}) = \text{E}\{\hat{\boldsymbol{\alpha}}_i(\boldsymbol{\theta}) - \tilde{\boldsymbol{\alpha}}_i(\boldsymbol{\theta})\}^2 = \text{tr}\{\mathbf{A}(\boldsymbol{\theta})\mathbf{B}(\boldsymbol{\theta})\},$$

with $\mathbf{A}(\boldsymbol{\theta}) = \text{Var}\{\mathbf{d}(\boldsymbol{\theta})\} = \{_m a_{rs}\}_{r,s=1,2}$, $\mathbf{d}(\boldsymbol{\theta}) = \frac{\partial\tilde{\alpha}_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = (d_1(\boldsymbol{\theta}), d_2(\boldsymbol{\theta}))^T$, and $\mathbf{B}(\boldsymbol{\theta})$ is either the MSE matrix $\text{E}\{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T\}$ or some approximation to the MSE matrix. In (4.4.1), We get

$$\begin{aligned} a_{11} &= \boldsymbol{\lambda}_i^T\{\mathbf{Z}^T\mathbf{P}(\boldsymbol{\theta}) - \mathbf{D}(\boldsymbol{\theta})\mathbf{Z}^T\mathbf{P}(\boldsymbol{\theta})\mathbf{Z}\mathbf{Z}^T\mathbf{P}(\boldsymbol{\theta})\}\mathbf{V}(\boldsymbol{\theta})\{\mathbf{P}(\boldsymbol{\theta})\mathbf{Z} - \mathbf{P}\mathbf{Z}\mathbf{Z}^T\mathbf{P}\mathbf{Z}\mathbf{D}(\boldsymbol{\theta})\}\boldsymbol{\lambda}_i, \\ a_{22} &= \boldsymbol{\lambda}_i^T\mathbf{D}(\boldsymbol{\theta})\mathbf{Z}^T\mathbf{P}(\boldsymbol{\theta})\mathbf{P}(\boldsymbol{\theta})\mathbf{V}(\boldsymbol{\theta})\mathbf{P}(\boldsymbol{\theta})\mathbf{P}(\boldsymbol{\theta})\mathbf{P}(\boldsymbol{\theta})\mathbf{Z}\mathbf{D}(\boldsymbol{\theta})\boldsymbol{\lambda}_i, \\ a_{12} &= a_{21} = -\boldsymbol{\lambda}_i^T\mathbf{D}(\boldsymbol{\theta})\mathbf{Z}^T\mathbf{P}(\boldsymbol{\theta})\mathbf{P}(\boldsymbol{\theta})\mathbf{V}(\boldsymbol{\theta})\{\mathbf{P}(\boldsymbol{\theta})\mathbf{Z} - \mathbf{P}(\boldsymbol{\theta})\mathbf{Z}\mathbf{Z}^T\mathbf{P}(\boldsymbol{\theta})\mathbf{Z}\mathbf{D}(\boldsymbol{\theta})\}\boldsymbol{\lambda}_i, \end{aligned}$$

and

$$\mathbf{B}(\boldsymbol{\theta}) = 2\begin{pmatrix} \text{tr}\{\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{Z}\mathbf{Z}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{Z}\mathbf{Z}^T\} & \text{tr}\{\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{Z}\mathbf{Z}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\} \\ \text{tr}\{\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{Z}\mathbf{Z}^T\} & \text{tr}\{\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{V}^{-1}(\boldsymbol{\theta})\} \end{pmatrix}^{-1}.$$

Prasad and Rao (1990) proposed the biased adjusted estimator shown as

$$\text{MSE}_{PR} = \text{E}\{\tilde{\boldsymbol{\alpha}}_i(\boldsymbol{\theta}) - \boldsymbol{\alpha}_i\}^2 + \text{E}\{\hat{\boldsymbol{\alpha}}_i(\boldsymbol{\theta}) - \tilde{\boldsymbol{\alpha}}_i(\boldsymbol{\theta})\}^2 = M_1(\boldsymbol{\theta}) + 2M_{PR_2}(\boldsymbol{\theta}). \tag{4.4.4}$$

In (4.4.4), we write

$$M_{PR_2}(\boldsymbol{\theta}) = \{\Phi(\boldsymbol{\theta})\mathbf{V}(\boldsymbol{\theta})\Phi^T(\boldsymbol{\theta})\mathbf{B}(\boldsymbol{\theta})\},$$

where $\rho(\boldsymbol{\theta}) = \boldsymbol{\lambda}_i^T \mathbf{D}(\boldsymbol{\theta})\mathbf{Z}^T\mathbf{V}^{-1}(\boldsymbol{\theta})$ and

$$\Phi(\boldsymbol{\theta}) = \mathrm{col}\{\frac{\partial\rho(\boldsymbol{\theta})}{\partial\sigma_\alpha^2}, \frac{\partial\rho(\boldsymbol{\theta})}{\partial\sigma_e^2}\} = \begin{pmatrix} \boldsymbol{\lambda}_i^T\mathbf{Z}^T\mathbf{V}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\lambda}_i^T\mathbf{D}(\boldsymbol{\theta})\mathbf{Z}^T\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{Z}\mathbf{Z}^T\mathbf{V}^{-1}(\boldsymbol{\theta}) \\ -\boldsymbol{\lambda}_i^T\mathbf{D}(\boldsymbol{\theta})\mathbf{Z}\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{V}^{-1}(\boldsymbol{\theta}) \end{pmatrix}.$$

For each simulation, we generated 3000 samples, fitted the model using REML and then computed the EBLUPs. For each of the balanced simulation settings, we computed the mean squared error MSE_T for $\hat{\alpha}_1$ as the average over the 3000 simulations of the squared difference between $\hat{\alpha}_1$ and the realized value α_1 as well as the averages of the estimates of the mean squared error estimates $\widehat{\mathrm{MSE}}_{LW,1}$, $\widehat{\mathrm{MSE}}_{KH,1}$ and $\widehat{\mathrm{MSE}}_{PR,i}$. Following Jiang et al. (2002), we report the average relative bias (ARB) defined as

$$ARB = \frac{\widehat{\mathrm{MSE}} - \mathrm{MSE}_T}{\mathrm{MSE}_T} \times 100\%,$$

where $\widehat{\mathrm{MSE}}$ denotes one of the three average mean squared error estimates. For the unbalanced simulation settings, we did the same calculations for $\hat{\alpha}_1$, $\hat{\alpha}_6$, $\hat{\alpha}_{11}$, $\hat{\alpha}_{16}$ and $\hat{\alpha}_{21}$ to capture $m = 10, 15, 25, 35, 40$ in the same simulation. We write the simulation codes in R (see Appendix)according to the following algorithm.

Table 4.1: Simulation setting: $\boldsymbol{\alpha}$ under normal distribution

Item	g	m	σ_α^2	σ_e^2	Item	g	m	σ_α^2	σ_e^2	Item	g	m	σ_α^2	σ_e^2	Item	g	m	σ_α^2	σ_e^2
(1)	10	10	25	49	(17)	10	10	81	49	(33)	10	10	25	96	(49)	10	10	81	96
(2)	10	25	25	49	(18)	10	25	81	49	(34)	10	25	25	96	(50)	10	25	81	96
(3)	10	50	25	49	(19)	10	50	81	49	(35)	10	50	25	96	(51)	10	50	81	96
(4)	10	100	25	49	(20)	10	100	81	49	(36)	10	100	25	96	(52)	10	100	81	96
(5)	25	10	25	49	(21)	25	10	81	49	(37)	25	10	25	96	(53)	25	10	81	96
(6)	25	25	25	49	(22)	25	25	81	49	(38)	25	25	25	96	(54)	25	25	81	96
(7)	25	50	25	49	(23)	25	50	81	49	(39)	25	50	25	96	(55)	25	50	81	96
(8)	25	100	25	49	(24)	25	100	81	49	(40)	25	100	25	96	(56)	25	100	81	96
(9)	50	10	25	49	(25)	50	10	81	49	(41)	50	10	25	96	(57)	50	10	81	96
(10)	50	25	25	49	(26)	50	25	81	49	(42)	50	25	25	96	(58)	50	25	81	96
(11)	50	50	25	49	(27)	50	50	81	49	(43)	50	50	25	96	(59)	50	50	81	96
(12)	50	100	25	49	(28)	50	100	81	49	(44)	50	100	25	96	(60)	50	100	81	96
(13)	100	10	25	49	(29)	100	10	81	49	(45)	100	10	25	96	(61)	100	10	81	96
(14)	100	25	25	49	(30)	100	25	81	49	(46)	100	25	25	96	(62)	100	25	81	96
(15)	100	50	25	49	(31)	100	50	81	49	(47)	100	50	25	96	(63)	100	50	81	96
(16)	100	100	25	49	(32)	100	100	81	49	(48)	100	100	25	96	(64)	100	100	81	96
(65), g=25,m=(10,10,10,10,10,15,15,15,15,15,25,25,25,25,25,35,35,35,35,35,40,40,40,40,40), $\sigma_\alpha^2 = 81$, $\sigma_e^2 = 49$																			

Table 4.2: Simulation setting: $\boldsymbol{\alpha}$ under mixture distribution

Item	g	m	σ_α^2	σ_e^2	Item	g	m	σ_α^2	σ_e^2	Item	g	m	σ_α^2	σ_e^2	Item	g	m	σ_α^2	σ_e^2
(1)	10	10	I	49	(17)	10	10	I	96	(33)	10	10	II	49	(49)	10	10	II	96
(2)	10	25	I	49	(18)	10	25	I	96	(34)	10	25	II	49	(50)	10	25	II	96
(3)	10	50	I	49	(19)	10	50	I	96	(35)	10	50	II	49	(51)	10	50	II	96
(4)	10	100	I	49	(20)	10	100	I	96	(36)	10	100	II	49	(52)	10	100	II	96
(5)	25	10	I	49	(21)	25	10	I	96	(37)	25	10	II	49	(53)	25	10	II	96
(6)	25	25	I	49	(22)	25	25	I	96	(38)	25	25	II	49	(54)	25	25	II	96
(7)	25	50	I	49	(23)	25	50	I	96	(39)	25	50	II	49	(55)	25	50	II	96
(8)	25	100	I	49	(24)	25	100	I	96	(40)	25	100	II	49	(56)	25	100	II	96
(9)	50	10	I	49	(25)	50	10	I	96	(41)	50	10	II	49	(57)	50	10	II	96
(10)	50	25	I	49	(26)	50	25	I	96	(42)	50	25	II	49	(58)	50	25	II	96
(11)	50	50	I	49	(27)	50	50	I	96	(43)	50	50	II	49	(59)	50	50	II	96
(12)	50	100	I	49	(28)	50	100	I	96	(44)	50	100	II	49	(60)	50	100	II	96
(13)	100	10	I	49	(29)	100	10	I	96	(45)	100	10	II	49	(61)	100	10	II	96
(14)	100	25	I	49	(30)	100	25	I	96	(46)	100	25	II	49	(62)	100	25	II	96
(15)	100	50	I	49	(31)	100	50	I	96	(47)	100	50	II	49	(63)	100	50	II	96
(16)	100	100	I	49	(32)	100	100	I	96	(48)	100	100	II	49	(64)	100	100	II	96
(65), g=25,m=(10,10,10,10,10,15,15,15,15,15,25,25,25,25,25,35,35,35,35,35,40,40,40,40,40), II $= 0.3N(14,9)+0.7N(-6,36)$, $\sigma_e^2 = 49$																			

NOTE: I $= 0.3N(14,25)+0.7N(-6,81)$; II $= 0.3N(14,9)+0.7N(-6,36)$

Algorithm 1 Simulation of different approximations to the MSE of the prediction

Require:

Let $t_i = \alpha_i$ and $\tilde{\boldsymbol{\tau}}_i(\boldsymbol{\theta}) = \boldsymbol{\lambda}_i^T \mathbf{D}(\boldsymbol{\theta})\mathbf{Z}^T\mathbf{P}(\boldsymbol{\theta})\mathbf{y}$ for $i = 1,\ldots,g$;
Set the parameters from Table 1 or Table 2;
for $r = 1$ to 3000 **do**
  Use equation (4.4.1) to generate observations $\mathbf{y}$;
  Obtain $\hat{\boldsymbol{\theta}}(\mathbf{y})$, the estimator of $\boldsymbol{\theta}$ by REML;
  Compute $\hat{\boldsymbol{\tau}}_i = \hat{\boldsymbol{\tau}}_i(\hat{\boldsymbol{\theta}})$ using $\hat{\boldsymbol{\theta}}(\mathbf{y})$;
  Obtain $\mathrm{MSE}_{LW}^{(r)}(i)$ as described in section (4.3.5),where (r) denotes the r-th time replications;
  Obtain $\mathrm{MSE}_{KH}^{(r)}(i)$ from Kackar's approximation (4.4.3);
  Obtain $\mathrm{MSE}_{PR}^{(r)}(i)$ from Prasad's approximation (4.4.4);
  Obtain $\mathrm{MSE}_{T}^{(r)}(i) = (\hat{\boldsymbol{\alpha}}_i - \boldsymbol{\alpha}_i)^2$
end for
$\mathrm{MSE}_T(i) = \frac{1}{3000}\sum_{r=1}^{3000} \mathrm{MSE}_T^{(r)}(i)$;
$\mathrm{MSE}_{LW}(i) = \frac{1}{3000}\sum_{r=1}^{3000} \mathrm{MSE}_{LW}^{(r)}(i)$;
$\mathrm{MSE}_{KH}(i) = \frac{1}{3000}\sum_{r=1}^{3000} \mathrm{MSE}_{KH}^{(r)}(i)$;
$\mathrm{MSE}_{PR}(i) = \frac{1}{3000}\sum_{r=1}^{3000} \mathrm{MSE}_{PR}^{(r)}(i)$;

4.4.2 Simulation results

After running the above algorithm in R, we obtain all simulation results listed in appendix.

4.5 Conclusion

Using simulation results, we draw the following graphs.

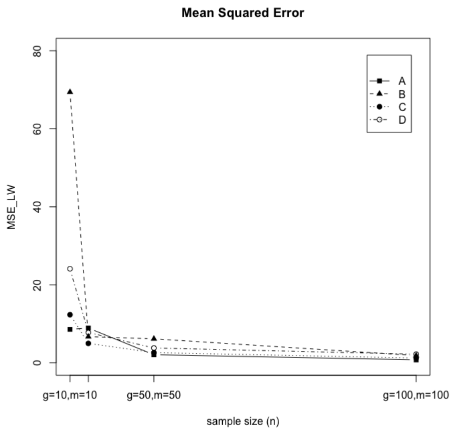

Figure 4.2: MSE_{LW} under Normal distribution with both the number of clusters and the cluster sizes growing simultaneously (The labels A, B, C and D for the curves represent the variance component combinations $(\sigma_\alpha^2, \sigma_e^2)$ equal to $(25, 49)$, $(81, 49)$, $(25, 96)$, $(81, 96)$, respectively)

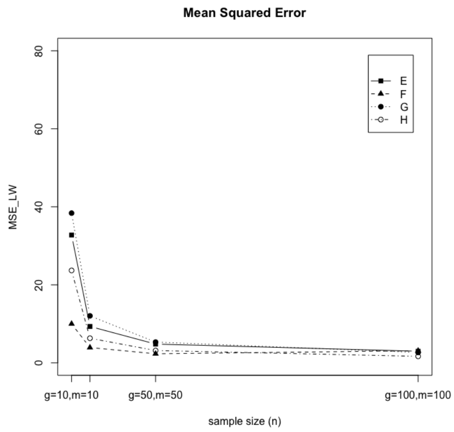

Figure 4.3: MSE_{LW} under Mixture distribution with both the number of clusters and the cluster sizes growing simultaneously (the labels E, F, G and H for the curves represent the variance component combinations $(\sigma_\alpha^2, \sigma_e^2)$ equal to $(148.2, 49)$, $(148.2, 96)$, $(111.9, 49)$, $(111.9, 96)$, respectively)

Fig.4.2 shows the relationship between MSE_{LW} and sample size (both the number of clusters and cluster sizes increasing simultaneously) under normality. We see that when both the number of clusters and cluster sizes are rising simultaneously, MSE_{LW} is declining regardless of different variance components. We find that MSE_{LW} will converge to zero if both numbers of clusters and cluster sizes tend to infinity. Fig.4.3 presents results under mixture distribution. We see that when both the number of clusters and cluster sizes are growing simultaneously, MSE_{LW} is decreasing regardless of different variance components. We also find that MSE_{LW} will converge to zero if both numbers of clusters and cluster sizes tend to infinity. Consequently, we can say when the number of clusters and cluster size large enough, our approximation MSE_{KH} converge to zero in general case.

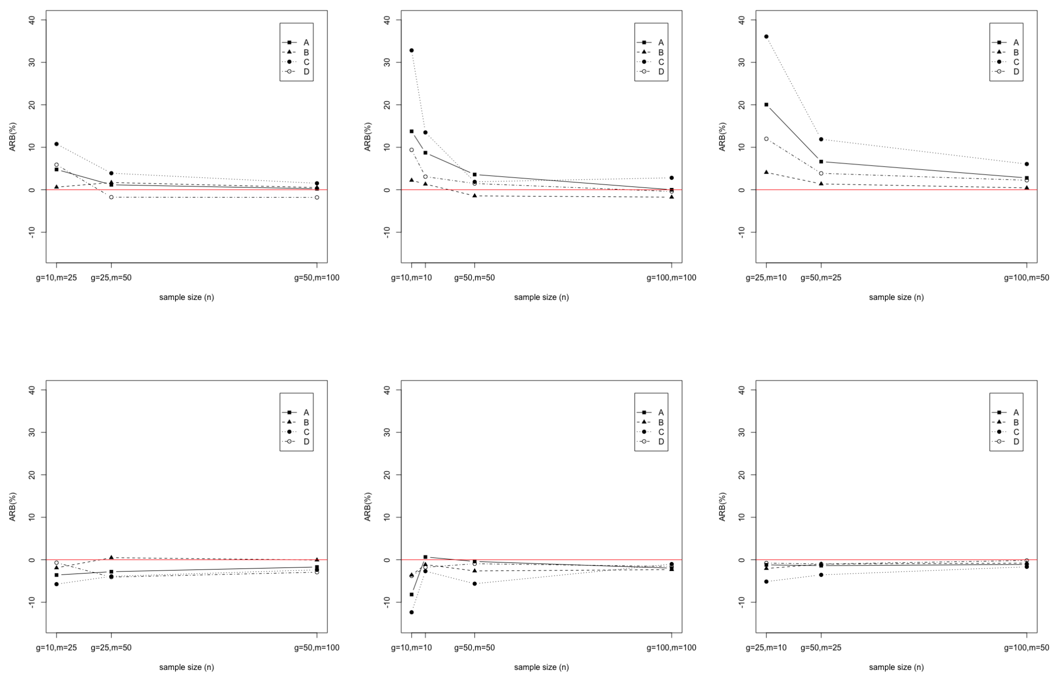

Figure 4.4: Average relative bias (ARB) of the mean squared error estimators under normal distributions with both the number of clusters and the cluster sizes growing simultaneously. The top row shows the ARB of $\widehat{\mathrm{MSE}}_{LW}$ and the second shows the ARB of $\widehat{\mathrm{MSE}}_{KH}$ and $\widehat{\mathrm{MSE}}_{PR}$. (These are so close together that they are represented by the same curves.) The three columns represent the cases $g/m \to 1/2$, $g/m = 1$, and $g/m \to 2$. The labels A, B, C and D for the curves within the sub-figures represent the variance component combinations $(\sigma_\alpha^2, \sigma_e^2)$ equal to $(25, 49)$, $(81, 49)$, $(25, 96)$, $(81, 96)$, respectively.

Figures 4.4 and 4.5 show the average relative bias of the mean squared error estimators under normal and non-normal distributions respectively, with both the number of clusters and the cluster sizes growing simultaneously. The top rows show the ARB of $\widehat{\mathrm{MSE}}_{LW}$ and the second rows show the ARB of $\widehat{\mathrm{MSE}}_{KH}$ and $\widehat{\mathrm{MSE}}_{PR}$. (These are so close together that they are represented by the same curves.) The three columns represent the cases $g/m \to 1/2$, $g/m = 1$ and $g/m \to 2$. The labels A, B, C and D for the curves within the sub-figures in Figures 4.4 represent the variance component combinations $(\sigma_\alpha^2, \sigma_e^2)$ equal to $(25, 49)$, $(81, 49)$, $(25, 96)$, $(81, 96)$, respectively; the labels E, F, G and H for the curves within the sub-figures in Figures 4.5 represent the variance component combinations $(\sigma_\alpha^2, \sigma_e^2)$ equal to $(148.2, 49)$, $(148.2, 96)$, $(111.9, 49)$, $(111.9, 96)$, respectively.

As expected in the balanced case, we have $\widehat{\mathrm{MSE}}_{KH,i} \le \widehat{\mathrm{MSE}}_{PR,i} \le \widehat{\mathrm{MSE}}_{LW,i}$. The estimates $\widehat{\mathrm{MSE}}_{KH,i} = \widehat{\mathrm{MSE}}_{PR,i}$ except in some very small samples ($g = m = 10$). In small samples, $\widehat{\mathrm{MSE}}_{LW}$ tends to be conservative (which is good) but performs well as g and m increase. The size of g and m at which good performance starts depends on the variance components and the relative sizes of g and m. If m is similar to or larger than g we get good results with quite small

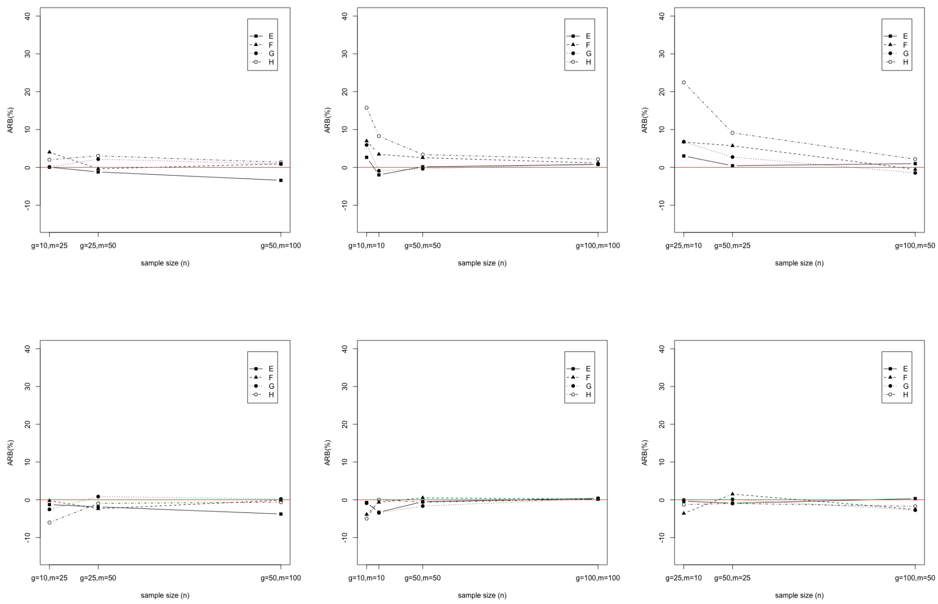

Figure 4.5: Average relative bias (ARB) of the mean squared error estimators under mixture distributions with both the number of clusters and the cluster sizes growing simultaneously. The top row shows the ARB of $\widehat{\mathrm{MSE}}_{LW}$ and the second shows the ARB of $\widehat{\mathrm{MSE}}_{KH}$ and $\widehat{\mathrm{MSE}}_{PR}$. (These are so close together that they are represented by the same curves.) The three columns represent the cases $g/m \to 1/2$, $g/m = 1$, and $g/m \to 2$. The labels E, F, G and H for the curves within the sub-figures represent the variance component combinations $(\sigma_\alpha^2, \sigma_e^2)$ equal to $(148.2, 49)$, $(111.9, 49)$, $(148.2, 96)$, $(111.9, 96)$, respectively.

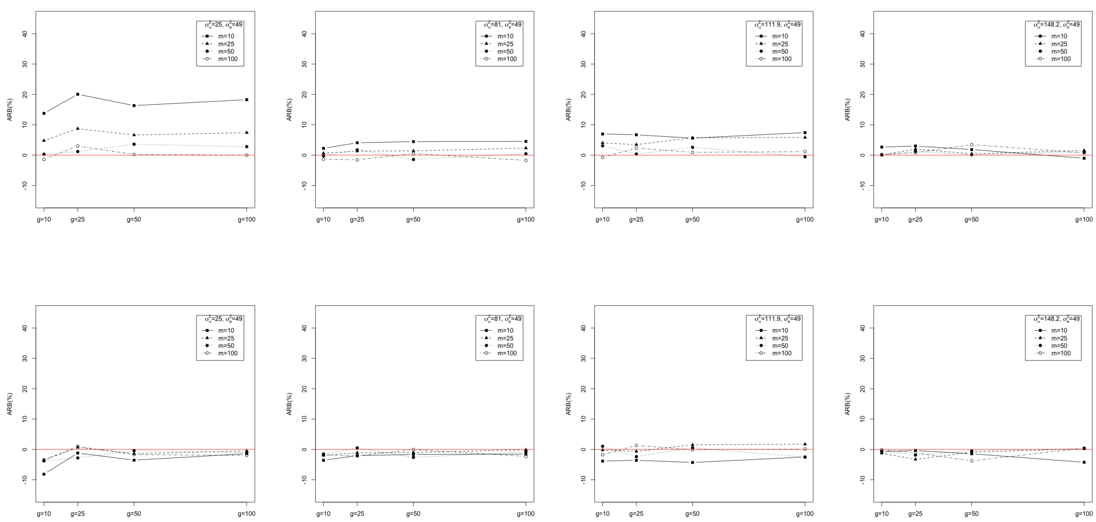

Figure 4.6: Average relative bias (ARB) of the mean squared error estimators under normal and mixture distributions showing the number of clusters growing with the cluster sizes fixed. The top row shows the ARB of $\widehat{\mathrm{MSE}}_{LW}$ and the second shows the ARB of $\widehat{\mathrm{MSE}}_{KH}$ and $\widehat{\mathrm{MSE}}_{PR}$. (These are so close together that they are represented by the same curves.) The first two columns represent normal distributions with variance component combinations $(\sigma_\alpha^2, \sigma_e^2)$ equal to $(25, 49)$ and $(81, 49)$; the last two columns represent mixture distributions with $(111.9, 49)$ and $(148.2, 49)$ respectively.

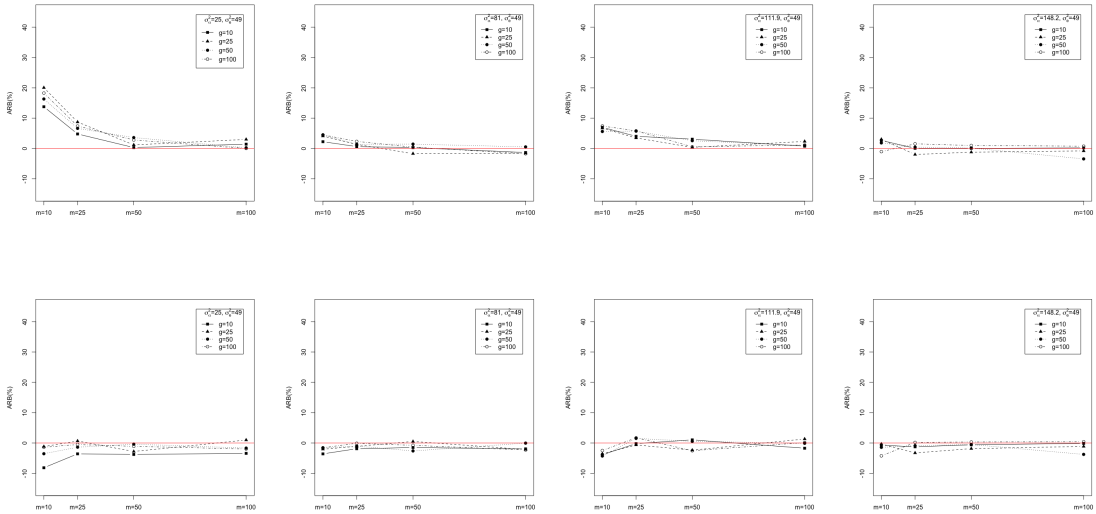

Figure 4.7: Average relative bias (ARB) of the mean squared error estimators under normal and mixture distributions showing the cluster size growing with the number of clusters fixed. The top row shows the ARB of $\widehat{\mathrm{MSE}}_{LW}$ and the second shows the ARB of $\widehat{\mathrm{MSE}}_{KH}$ and $\widehat{\mathrm{MSE}}_{PR}$. (These are so close together that they are represented by the same curves.) The first two columns represent normal distributions with variance component combinations $(\sigma_\alpha^2, \sigma_e^2)$ equal to $(25, 49)$ and $(81, 49)$; the last two columns represent mixture distributions with $(111.9, 49)$ and $(148.2, 49)$ respectively.

samples (e.g. $g = m = 50$ or $g = 25$ and $m = 50$) but, if g is larger than m, we require larger samples for good performance (e.g. $g = 100$ and $m = 50$). Nonetheless, in many problems these are relatively small sample sizes. It is interesting that $\widehat{\mathrm{MSE}}_{LW}$ performs better for larger σ_α^2. For comparison, $\widehat{\mathrm{MSE}}_{KH}$ and $\widehat{\mathrm{MSE}}_{PR}$ tend to be slightly optimistic in very small samples and then also perform increasingly well as g and m increase. Figure 4.5 shows that the results are quite similar for the mixture case.

Figures 4.4 and 4.5 confirm the theoretical results of the paper and show the estimators working well. It is interesting to also consider cases where the conditions for the theoretical results are not met. We present results showing how the estimators perform when the number of clusters grows but the cluster size is fixed (Figure 4.6) and when the number of clusters is fixed while the cluster size grows (Figure 4.7).

We see from Figure 4.6 that when g increases with m held fixed, the ARB of $\widehat{\mathrm{MSE}}_{LW}$ tends to be constant for both the normal and mixture distributions. The results are generally better as m increases within each sub-figure and when σ_α^2 is larger. The results for $\widehat{\mathrm{MSE}}_{KH}$ and $\widehat{\mathrm{MSE}}_{PR}$ are better than those for $\widehat{\mathrm{MSE}}_{LW}$ but still show a tendency to be slightly optimistic in small samples. Figure 4.7 shows that $\widehat{\mathrm{MSE}}_{LW}$ performs very well as the cluster size m increases when the number of clusters g is fixed. This confirms that the size of m is somewhat more important than the size of g for our approximation to perform well.

Chapter 5 Conclusion and further research

5.1 Conclusion

In Chapter 2, we obtained the asymptotic properties of ML and REML estimators in the nested error regression model with balanced data ($m_i = m$) under mild conditions, e.g., assuming non-normality, and both the number of clusters and cluster size go to infinity. A breakthrough is distinguishing a between cluster covariate that is constant within clusters $x_{ij} = x_i$ from a within cluster covariate that varies within clusters. The results show explicitly that the between and within parameters are estimated at different rates and are asymptotically orthogonal. This means that within cluster covariates should have their between cluster contribution (as represented by the within cluster means) separated out and included in the model as contextual effects. We explained the conditions that are needed and described the form of the asymptotic variance of the estimators that we obtain.

Chapter 3 extends the results of Chapter 2 to unbalanced data. This extension is by no means trivial because, in establish a uniform asymptotic linear approximation to the ML estimating equations that hold in an appropriate shrinking neighborhood of the true parameters, and then obtain consistency and asymptotic normality for the ML estimators from the uniform asymptotic linear approximation. To establish the asymptotic properties of the REML estimator, we use a profile function argument instead of working directly with the structure of the estimating equation as we did in the balanced case. We have constructed an adjusted maximum likelihood (AML) estimating equation with the additional term $\log |\mathbf{X}^T \mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{X}|$, and it shows that it has same asymptotic properties obtained from the log-likelihood with ML estimators. Then we use a profile function approach to show that the REML estimating equation is the profile version of the adjusted maximum likelihood estimating function. Finally, we obtain the asymptotic properties of REML estimators in general (non-normal) nested regression model with unbalanced data, when both the number of clusters and cluster size go to infinity.

In Chapter 4, we have used these results to describe the asymptotic behavior of EBLUPs of the unobserved random random effects in the nested error regression model. In particular, we obtained the asymptotic distribution of EBLUPs and showed that the distribution of EBLUPs can be approximated by the convolution of the true distribution of the random effects and a normal distribution which depends on the number of clusters and the cluster size. This simple new result gives considerable insight into the asymptotic behavior of EBPLUPs. Our approximation and estimator of the mean squared error of EBLUPs works well even when m is small. And ur approximation and estimator of the mean squared error of EBLUPs is much simpler than the well established versions of Kackar and Harville (1984) and Prasad and Rao (1990). These versions were developed under the asymptotic framework in which the number of clusters goes to infinity with the clusters sizes fixed. Since the EBLUPs are not consistent under this framework, Kackar and Harville (1984) and Prasad and Rao (1990) were forced to approximate the mean squared error of the EBLUPS by taking the mean squared error and then letting the number of clusters go to infinity. We have taken the other approach of working in an asymptotic framework in which the EBLUPs are consistent, making a linear approximation to the EBLUPs and then taking the mean squared error. That is, the order of computing the mean squared error and taking limits is different. The Kackar and Harville (1984) and Prasad and Rao (1990) approximations and estimators converge to ours when the cluster sizes also increase to infinity; the advantage of our approach is its simplicity (in conditions and in execution). Also, in our approach BLUPs and EBLUPs have the same asymptotic distribution, a result that fits BLUPs and EBLUPs into the same situation as methods like generalized least squares for estimating regression parameters where the estimators have the same asymptotic distribution when the variance parameters are known and when they are estimated consistently. Finally, we presented both a simulation study to show the good performance of our approximation and its estimator in finite samples. We have worked with a general regression model but with the simplest possible random effect structure. In spite of it simplicity, the model is nonetheless useful and is used in practice. The advantage of considering a simple model is that we can gain considerable insight into the properties of EBLUPs, their asymptotic distributions and mean squared errors. This has enabled us to clarify the relationship between the Kackar and Harville (1984) and Prasad and Rao (1990) approximations and estimators to each other and to our approximations and estimators. The success of our approach and the insights it has provided in the simple nested error regression model justify our generalizing the results to models with more complicated random structures.

5.2 Further research

In this thesis, we studied the asymptotic theories of ML and REML estimators and asymptotic properties of EBLUPs in the nested error regression model. Based on these research results, we plan to explore the EBLUPs associated with a mixed effect, which is a function of fixed and random effects. For example, in small area estimation (SAE), the interest is typically about the small area means or proportions, which are mixed effects under a mixed effect model. In addition to this, we will study some potential applications of the theory/methods developed in this thesis with some real-data example(s). In the simulation study, as the MSPE estimation is first-order unbiased, we will study jackknife and/or bootstrap methods for the MSPE estimation. Moreover, we plan to extend the nested error regression model to the two-way crossed models, even more general linear mixed models.

Appendix A

We here give the proof of lemma 1.

Proof.

$$\sum_{j=1}^{m}\sum_{k=1}^{m}\sum_{r=1}^{m}\sum_{s=1}^{m}\mathrm{E}(e_{ij}e_{ik}e_{ir}e_{is})$$

$$= \sum_{j=1}^{m}\sum_{r=1}^{m}\mathrm{E}(e_{ij}^2e_{ir}^2) + \sum_{j=1}^{m}\sum_{k\neq j}^{m}\sum_{r=1}^{m}\mathrm{E}(e_{ij}e_{ik}e_{ir}^2) + \sum_{j=1}^{m}\sum_{k=1}^{m}\sum_{r=1}^{m}\sum_{s\neq r}^{m}\mathrm{E}(e_{ij}e_{ik}e_{ir}e_{is})$$

$$= \sum_{j=1}^{m}\sum_{r=1}^{m}\mathrm{E}(e_{ij}^2e_{ir}^2) + \sum_{j=1}^{m}\sum_{k=1}^{m}\sum_{r=1}^{m}\sum_{s\neq r}^{m}\mathrm{E}(e_{ij}e_{ik}e_{ir}e_{is})$$

$$= \sum_{j=1}^{m}\sum_{r=1}^{m}\mathrm{E}(e_{ij}^2e_{ir}^2) + \sum_{j=1}^{m}\sum_{r=1}^{m}\sum_{s\neq r}^{m}\mathrm{E}(e_{ij}^2e_{ir}e_{is}) + \sum_{j=1}^{m}\sum_{k\neq j}^{m}\sum_{r=1}^{m}\sum_{s\neq r}^{m}\mathrm{E}(e_{ij}e_{ik}e_{ir}e_{is})$$

$$= \sum_{j=1}^{m}\sum_{r=1}^{m}\mathrm{E}(e_{ij}^2e_{ir}^2) + \sum_{j=1}^{m}\sum_{k\neq j}^{m}\sum_{r=1}^{m}\sum_{s\neq r}^{m}\mathrm{E}(e_{ij}e_{ik}e_{ir}e_{is})$$

$$= \sum_{j=1}^{m}\mathrm{E}(e_{ij}^4) + \sum_{j=1}^{m}\sum_{r\neq j}^{m}\mathrm{E}(e_{ij}^2)\,\mathrm{E}(e_{ir}^2) + \sum_{j=1}^{m}\sum_{k\neq j}^{m}\sum_{r=1}^{m}\sum_{s\neq r}^{m}\mathrm{E}(e_{ij}e_{ik}e_{ir}e_{is})$$

$$= \sum_{j=1}^{m}\mathrm{E}(e_{ij}^4) + m(m-1)\dot{\sigma_e}^4 + \sum_{j=1}^{m}\sum_{k\neq j}^{m}\sum_{r=1}^{m}\sum_{s\neq r}^{m}\mathrm{E}(e_{ij}e_{ik}e_{ir}e_{is})$$

$$= \sum_{j=1}^{m}\mathrm{E}(e_{ij}^4) + m(m-1)\dot{\sigma_e}^4 + \sum_{j=r=1}^{m}\sum_{k\neq j}^{m}\sum_{s\neq j}^{m}\mathrm{E}(e_{ij}^2)\,\mathrm{E}(e_{ik}e_{is}) + \sum_{j=1}^{m}\sum_{k\neq j}^{m}\sum_{r\neq j}^{m}\sum_{s\neq r}^{m}\mathrm{E}(e_{ij}e_{ik}e_{ir}e_{is})$$

$$= \sum_{j=1}^{m}\mathrm{E}(e_{ij}^4) + m(m-1)\dot{\sigma_e}^4 + \sum_{j=1}^{m}\sum_{\substack{k\neq j\\ k=s}}^{m}\mathrm{E}(e_{ij}^2)\,\mathrm{E}(e_{ik}^2) + \sum_{j=1}^{m}\sum_{k\neq j}^{m}\sum_{\substack{s\neq j\\ s\neq k}}^{m}\mathrm{E}(e_{ij}^2)\,\mathrm{E}(e_{ik})\,\mathrm{E}(e_{is})$$

$$+ \sum_{j=1}^{m}\sum_{k\neq j}^{m}\sum_{r\neq j}^{m}\sum_{s\neq r}^{m}\mathrm{E}(e_{ij}e_{ik}e_{ir}e_{is})$$

$$= \sum_{j=1}^{m}\mathrm{E}(e_{ij}^4) + m(m-1)\dot{\sigma_e}^4 + m(m-1)\dot{\sigma_e}^4 + \sum_{j=1}^{m}\sum_{k\neq j}^{m}\sum_{\substack{s\neq j\\ s\neq k}}^{m}\mathrm{E}(e_{ij}^2)\,\mathrm{E}(e_{ik})\,\mathrm{E}(e_{is}) + \sum_{j=1}^{m}\sum_{k\neq j}^{m}\sum_{r\neq j}^{m}\sum_{s\neq r}^{m}\mathrm{E}(e_{ij}e_{ik}e_{ir}e_{is})$$

$$= \sum_{j=1}^{m}\mathrm{E}(e_{ij}^4) + 2m(m-1)\dot{\sigma_e}^4 + \sum_{j=1}^{m}\sum_{k\neq j}^{m}\sum_{r\neq j}^{m}\sum_{s\neq r}^{m}\mathrm{E}(e_{ij}e_{ik}e_{ir}e_{is})$$

$$= \sum_{j=1}^{m}\mathrm{E}(e_{ij}^4) + 2m(m-1)\dot{\sigma_e}^4 + \sum_{j=1}^{m}\sum_{\substack{k\neq j\\ k=r}}^{m}\sum_{s\neq k}^{m}\mathrm{E}\,e_{ij}e_{ik}^2e_{is} + \sum_{j=1}^{m}\sum_{k\neq j}^{m}\sum_{\substack{r\neq j\\ r\neq k}}^{m}\sum_{s\neq r}^{m}\mathrm{E}(e_{ij}e_{ik}e_{ir}e_{is})$$

$$= \sum_{j=1}^{m}\mathrm{E}(e_{ij}^4) + 2m(m-1)\dot{\sigma_e}^4 + \sum_{j=1}^{m}\sum_{\substack{k\neq j\\ k=r}}^{m}\sum_{s\neq k}^{m}\mathrm{E}\,e_{ij}e_{ik}^2e_{is}$$

$$= \sum_{j=1}^{m}\mathrm{E}(e_{ij}^4) + 2m(m-1)\dot{\sigma_e}^4 + \sum_{j=1}^{m}\sum_{k\neq j}^{m}\sum_{\substack{s\neq k\\ s=j}}^{m}\mathrm{E}(e_{ij}^2)\,\mathrm{E}(e_{ik}^2) + \sum_{j=1}^{m}\sum_{k\neq j}^{m}\sum_{\substack{s\neq k\\ s\neq j}}^{m}\mathrm{E}(e_{ij})\,\mathrm{E}(e_{ik}^2)\,\mathrm{E}(e_{is})$$

$$= m\,\mathrm{E}(e_{ij}^4) + 3m(m-1)\dot{\sigma_e}^4.$$

□

Simulation results

Table A.1: Simulation results with normal distribution $(\sigma_\alpha^2 = 25, \sigma_e^2 = 49)$

(1). g=10, m=10			(2). g=10, m=25			(3). g=10, m=50			(4). g=10, m=100		
MSE_T	7.546	ARB	MSE_T	4.937	ARB	MSE_T	3.72	ARB	MSE_T	3.94	ARB
MSE_LW	8.584	13.76%	MSE_LW	5.173	4.78%	MSE_LW	3.731	0.30%	MSE_LW	3.884	-1.42%
MSE_KH	6.929	-8.18%	MSE_KH	4.76	-3.59%	MSE_KH	3.579	-3.79%	MSE_KH	3.805	-3.43%
MSE_PR	6.935	-8.10%	MSE_PR	4.761	-3.56%	MSE_PR	3.579	-3.79%	MSE_PR	3.805	-3.43%
(5). g=25, m=10			(6). g=25 m=25			(7). g=25, m=50			(8). g=25, m=100		
MSE_T	5.672	ARB	MSE_T	8.197	ARB	MSE_T	1.958	ARB	MSE_T	1.44	ARB
MSE_LW	6.809	20.05%	MSE_LW	8.912	8.72%	MSE_LW	1.981	1.17%	MSE_LW	1.483	2.99%
MSE_KH	5.604	-1.20%	MSE_KH	8.25	0.65%	MSE_KH	1.903	-2.81%	MSE_KH	1.454	0.97%
MSE_PR	5.606	-1.16%	MSE_PR	8.25	0.65%	MSE_PR	1.903	-2.81%	MSE_PR	1.454	0.97%
(9). g=50, m=10			(10). g=50, m=25			(11). g=50, m=50			(12). g=50, m=100		
MSE_T	4.652	ARB	MSE_T	2.308	ARB	MSE_T	2.005	ARB	MSE_T	1.706	ARB
MSE_LW	5.411	16.32%	MSE_LW	2.461	6.63%	MSE_LW	2.077	3.59%	MSE_LW	1.71	0.23%
MSE_KH	4.487	-3.55%	MSE_KH	2.276	-1.39%	MSE_KH	1.997	-0.40%	MSE_KH	1.677	-1.70%
MSE_PR	4.488	-3.53%	MSE_PR	2.276	-1.39%	MSE_PR	1.997	-0.40%	MSE_PR	1.677	-1.70%
(13). g=100, m=10			(14). g=100, m=25			(15). g=100, m=50			(16). g=100, m=100		
MSE_T	4.597	ARB	MSE_T	2.704	ARB	MSE_T	1.44	ARB	MSE_T	0.754	ARB
MSE_LW	5.437	18.27%	MSE_LW	2.904	7.40%	MSE_LW	1.4802	2.79%	MSE_LW	0.754	0.00%
MSE_KH	4.532	-1.41%	MSE_KH	2.69	-0.52%	MSE_KH	1.424	-1.11%	MSE_KH	0.739	-1.99%
MSE_PR	4.532	-1.41%	MSE_PR	2.69	-0.52%	MSE_PR	1.424	-1.11%	MSE_PR	0.739	-1.99%

Table A.2: Simulation results with normal distribution $(\sigma_\alpha^2 = 81, \sigma_e^2 = 49)$

(17). g=10, m=10			(18). g=10, m=25			(19). g=10, m=50			(20). g=10, m=100		
MSE_T	67.884	ARB	MSE_T	9.986	ARB	MSE_T	9.268	ARB	MSE_T	9.574	ARB
MSE_LW	69.401	2.23%	MSE_LW	10.048	0.62%	MSE_LW	9.242	-0.28%	MSE_LW	9.446	-1.34%
MSE_KH	65.44	-3.60%	MSE_KH	9.796	-1.90%	MSE_KH	9.128	-1.51%	MSE_KH	9.387	-1.95%
MSE_PR	65.443	-3.60%	MSE_PR	9.796	-1.90%	MSE_PR	9.128	-1.51%	MSE_PR	9.387	-1.95%
(21). g=25, m=10			(22). g=25 m=25			(23). g=25, m=50			(24). g=25, m=100		
MSE_T	10.802	ARB	MSE_T	6.619	ARB	MSE_T	5.491	ARB	MSE_T	5.906	ARB
MSE_LW	11.243	4.08%	MSE_LW	6.707	1.33%	MSE_LW	5.585	1.71%	MSE_LW	5.814	-1.56%
MSE_KH	10.581	-2.05%	MSE_KH	6.544	-1.13%	MSE_KH	5.518	0.49%	MSE_KH	5.779	-2.15%
MSE_PR	10.581	-2.05%	MSE_PR	6.544	-1.13%	MSE_PR	5.518	0.49%	MSE_PR	5.779	-2.15%
(25). g=50, m=10			(26). g=50, m=25			(27). g=50, m=50			(28). g=50, m=100		
MSE_T	6.808	ARB	MSE_T	7.103	ARB	MSE_T	6.245	ARB	MSE_T	2.107	ARB
MSE_LW	7.111	4.45%	MSE_LW	7.201	1.38%	MSE_LW	6.155	-1.44%	MSE_LW	2.118	0.52%
MSE_KH	6.694	-1.67%	MSE_KH	7.03	-1.03%	MSE_KH	6.081	-2.63%	MSE_KH	2.106	-0.05%
MSE_PR	6.694	-1.67%	MSE_PR	7.03	-1.03%	MSE_PR	6.081	-2.63%	MSE_PR	2.106	-0.05%
(29). g=100, m=10			(30). g=100, m=25			(31). g=100, m=50			(32). g=100, m=100		
MSE_T	6.362	ARB	MSE_T	2.761	ARB	MSE_T	2.046	ARB	MSE_T	1.9	ARB
MSE_LW	6.649	4.51%	MSE_LW	2.825	2.32%	MSE_LW	2.055	0.44%	MSE_LW	1.867	-1.74%
MSE_KH	6.263	-1.56%	MSE_KH	2.758	-0.11%	MSE_KH	2.03	-0.78%	MSE_KH	1.856	-2.32%
MSE_PR	6.263	-1.56%	MSE_PR	2.758	-0.11%	MSE_PR	2.03	-0.78%	MSE_PR	1.856	-2.32%

Table A.3: Simulation results with normal distribution $(\sigma_\alpha^2 = 25, \sigma_e^2 = 96)$

(33). g=10, m=10			(34). g=10, m=25			(35). g=10, m=50			(36). g=10, m=100		
MSE_T	9.286	ARB	MSE_T	7.925	ARB	MSE_T	4.51	ARB	MSE_T	8.208	ARB
MSE_LW	12.334	32.82%	MSE_LW	8.779	10.78%	MSE_LW	4.845	7.43%	MSE_LW	8.324	1.41%
MSE_KH	8.138	-12.36%	MSE_KH	7.47	-5.74%	MSE_KH	4.457	-1.18%	MSE_KH	8.006	-2.46%
MSE_PR	8.146	-12.28%	MSE_PR	7.471	-5.73%	MSE_PR	4.457	-1.18%	MSE_PR	8.006	-2.46%
(37). g=25, m=10			(38). g=25 m=25			(39). g=25, m=50			(40). g=25, m=100		
MSE_T	7.803	ARB	MSE_T	4.392	ARB	MSE_T	3.409	ARB	MSE_T	2.195	ARB
MSE_LW	10.618	36.08%	MSE_LW	4.985	13.50%	MSE_LW	3.542	3.90%	MSE_LW	2.278	3.78%
MSE_KH	7.402	-5.14%	MSE_KH	4.274	-2.69%	MSE_KH	3.276	-3.90%	MSE_KH	2.19	-0.23%
MSE_PR	7.402	-5.14%	MSE_PR	4.274	-2.69%	MSE_PR	3.277	-3.87%	MSE_PR	2.19	-0.23%
(41). g=50, m=10			(42). g=50, m=25			(43). g=50, m=50			(44). g=50, m=100		
MSE_T	7.864	ARB	MSE_T	4.087	ARB	MSE_T	2.572	ARB	MSE_T	1.434	ARB
MSE_LW	10.397	32.21%	MSE_LW	4.573	11.89%	MSE_LW	2.62	1.87%	MSE_LW	1.456	1.53%
MSE_KH	7.37	-6.28%	MSE_KH	3.942	-3.55%	MSE_KH	2.427	-5.64%	MSE_KH	1.4	-2.37%
MSE_PR	7.372	-6.26%	MSE_PR	3.942	-3.55%	MSE_PR	2.427	-5.64%	MSE_PR	1.4	-2.37%
(45). g=100, m=10			(46). g=100, m=25			(47). g=100, m=50			(48). g=100, m=100		
MSE_T	7.826	ARB	MSE_T	4.605	ARB	MSE_T	2.046	ARB	MSE_T	1.191	ARB
MSE_LW	10.457	33.62%	MSE_LW	5.258	14.18%	MSE_LW	2.17	6.06%	MSE_LW	1.2246	2.82%
MSE_KH	7.501	-4.15%	MSE_KH	4.549	-1.22%	MSE_KH	2.012	-1.66%	MSE_KH	1.179	-1.01%
MSE_PR	7.502	-4.14%	MSE_PR	4.549	-1.22%	MSE_PR	2.012	-1.66%	MSE_PR	1.179	-1.01%

Table A.4: Simulation results with normal distribution $(\sigma_\alpha^2 = 81, \sigma_e^2 = 96)$

(49). g=10, m=10			(50). g=10, m=25			(51). g=10, m=50			(52). g=10, m=100		
MSE_T	22.037	ARB	MSE_T	12.542	ARB	MSE_T	14.927	ARB	MSE_T	44.617	ARB
MSE_LW	24.106	9.39%	MSE_LW	13.282	5.90%	MSE_LW	14.905	-0.15%	MSE_LW	44.72	0.23%
MSE_KH	21.199	-3.80%	MSE_KH	12.629	0.69%	MSE_KH	14.545	-2.56%	MSE_KH	44.188	-0.96%
MSE_PR	21.203	-3.78%	MSE_PR	12.629	0.69%	MSE_PR	14.545	-2.56%	MSE_PR	44.188	-0.96%
(53). g=25, m=10			(54). g=25 m=25			(55). g=25, m=50			(56). g=25, m=100		
MSE_T	11.436	ARB	MSE_T	7.535	ARB	MSE_T	7.629	ARB	MSE_T	4.2	ARB
MSE_LW	12.808	12.00%	MSE_LW	7.768	3.09%	MSE_LW	7.496	-1.74%	MSE_LW	4.174	-0.62%
MSE_KH	11.348	-0.77%	MSE_KH	7.402	-1.77%	MSE_KH	7.318	-4.08%	MSE_KH	4.123	-1.83%
MSE_PR	11.349	-0.76%	MSE_PR	7.402	-1.77%	MSE_PR	7.318	-4.08%	MSE_PR	4.123	-1.83%
(57). g=50, m=10			(58). g=50, m=25			(59). g=50, m=50			(60). g=50, m=100		
MSE_T	10.993	ARB	MSE_T	5.324	ARB	MSE_T	3.752	ARB	MSE_T	3.282	ARB
MSE_LW	11.8	7.34%	MSE_LW	5.53	3.87%	MSE_LW	3.807	1.47%	MSE_LW	3.223	-1.80%
MSE_KH	10.502	-4.47%	MSE_KH	5.271	-1.00%	MSE_KH	3.717	-0.93%	MSE_KH	3.186	-2.93%
MSE_PR	10.503	-4.46%	MSE_PR	5.271	-1.00%	MSE_PR	3.717	-0.93%	MSE_PR	3.186	-2.93%
(61). g=100, m=10			(62). g=100, m=25			(63). g=100, m=50			(64). g=100, m=100		
MSE_T	9.893	ARB	MSE_T	5.105	ARB	MSE_T	2.986	ARB	MSE_T	2.229	ARB
MSE_LW	10.829	9.46%	MSE_LW	5.199	1.84%	MSE_LW	3.053	2.24%	MSE_LW	2.22	0.40%
MSE_KH	9.661	-2.35%	MSE_KH	4.96	-2.84%	MSE_KH	2.981	-0.17%	MSE_KH	2.194	1.57%
MSE_PR	9.661	-2.35%	MSE_PR	4.96	-2.84%	MSE_PR	2.981	-0.17%	MSE_PR	2.194	1.57%

Table A.5: Simulation results Under mixture distribution

(1). g=10, m=10			(2). g=10, m=25			(3). g=10, m=50			(4). g=10, m=100		
MSE_T	31.917	ARB	MSE_T	42.576	ARB	MSE_T	42.05	ARB	MSE_T	40.859	ARB
MSE_LW	32.763	2.65%	MSE_LW	42.598	0.05%	MSE_LW	42.069	0.05%	MSE_LW	40.935	0.19%
MSE_KH	31.686	-0.72%	MSE_KH	42.034	-1.27%	MSE_KH	41.793	-0.61%	MSE_KH	40.801	-0.14%
MSE_PR	31.687	-0.72%	MSE_PR	42.034	-1.27%	MSE_PR	41.793	-0.61%	MSE_PR	40.801	-0.14%
(5). g=25, m=10			(6). g=25 m=25			(7). g=25, m=50			(8). g=25, m=100		
MSE_T	12.986	ARB	MSE_T	9.517	ARB	MSE_T	10.369	ARB	MSE_T	6.919	ARB
MSE_LW	13.377	3.01%	MSE_LW	9.328	-1.99%	MSE_LW	10.242	-1.22%	MSE_LW	6.863	-0.81%
MSE_KH	12.939	-0.36%	MSE_KH	9.204	-3.29%	MSE_KH	10.175	-1.87%	MSE_KH	6.84	-1.14%
MSE_PR	12.939	-0.36%	MSE_PR	9.204	-3.29%	MSE_PR	10.175	-1.87%	MSE_PR	6.84	-1.14%
(9). g=50, m=10			(10). g=50, m=25			(11). g=50, m=50			(12). g=50, m=100		
MSE_T	9.215	ARB	MSE_T	5.643	ARB	MSE_T	4.797	ARB	MSE_T	6.344	ARB
MSE_LW	9.384	1.83%	MSE_LW	5.667	0.43%	MSE_LW	4.804	0.15%	MSE_LW	6.126	-3.44%
MSE_KH	9.081	-1.45%	MSE_KH	5.592	-0.90%	MSE_KH	4.773	-0.50%	MSE_KH	6.105	-3.77%
MSE_PR	9.081	-1.45%	MSE_PR	5.592	-0.90%	MSE_PR	4.773	-0.50%	MSE_PR	6.105	-3.77%
(13). g=100, m=10			(14). g=100, m=25			(15). g=100, m=50			(16). g=100, m=100		
MSE_T	6.446	ARB	MSE_T	4.442	ARB	MSE_T	7.275	ARB	MSE_T	2.942	ARB
MSE_LW	6.379	-1.04%	MSE_LW	4.51	1.53%	MSE_LW	7.346	0.98%	MSE_LW	2.964	0.75%
MSE_KH	6.172	-4.25%	MSE_KH	4.45	0.18%	MSE_KH	7.298	0.32%	MSE_KH	2.954	0.41%
MSE_PR	6.172	-4.25%	MSE_PR	4.45	0.18%	MSE_PR	7.298	0.32%	MSE_PR	2.954	0.41%

NOTE: I $= 0.3N(14, 25) + 0.7N(-6, 81)$; $\sigma_e^2 = 49$

Table A.6: Simulation results Under mixture distribution

(17). g=10, m=10			(18). g=10, m=25			(19). g=10, m=50			(20). g=10, m=100		
MSE_T	36.237	ARB	MSE_T	22.125	ARB	MSE_T	22.849	ARB	MSE_T	96.69	ARB
MSE_LW	38.384	5.92%	MSE_LW	22.148	0.10%	MSE_LW	22.998	0.65%	MSE_LW	97.525	0.86%
MSE_KH	35.916	-0.89%	MSE_KH	21.564	-2.54%	MSE_KH	22.7	-0.65%	MSE_KH	96.89	0.21%
MSE_PR	35.918	-0.88%	MSE_PR	21.564	-2.54%	MSE_PR	22.7	-0.65%	MSE_PR	96.89	0.21%
(21). g=25, m=10			(22). g=25 m=25			(23). g=25, m=50			(24). g=25, m=100		
MSE_T	14.749	ARB	MSE_T	12.149	ARB	MSE_T	9.019	ARB	MSE_T	8.241	ARB
MSE_LW	15.747	6.77%	MSE_LW	12.039	-0.91%	MSE_LW	9.215	2.17%	MSE_LW	8.009	-2.82%
MSE_KH	14.741	-0.05%	MSE_KH	11.73	-3.45%	MSE_KH	9.096	0.85%	MSE_KH	7.957	-3.45%
MSE_PR	14.741	-0.05%	MSE_PR	11.73	-3.45%	MSE_PR	9.096	0.85%	MSE_PR	7.957	-3.45%
(25). g=50, m=10			(26). g=50, m=25			(27). g=50, m=50			(28). g=50, m=100		
MSE_T	15.279	ARB	MSE_T	8.104	ARB	MSE_T	5.34	ARB	MSE_T	20.794	ARB
MSE_LW	16.072	5.19%	MSE_LW	8.325	2.73%	MSE_LW	5.32	-0.37%	MSE_LW	20.974	0.87%
MSE_KH	15.074	-1.34%	MSE_KH	8.112	0.10%	MSE_KH	5.252	-1.65%	MSE_KH	20.839	0.22%
MSE_PR	15.074	-1.34%	MSE_PR	8.112	0.10%	MSE_PR	5.252	-1.65%	MSE_PR	20.839	0.22%
(29). g=100, m=10			(30). g=100, m=25			(31). g=100, m=50			(32). g=100, m=100		
MSE_T	13.858	ARB	MSE_T	5.402	ARB	MSE_T	3.557	ARB	MSE_T	2.635	ARB
MSE_LW	14.448	4.26%	MSE_LW	5.366	-0.67%	MSE_LW	3.505	-1.46%	MSE_LW	2.66	0.95%
MSE_KH	13.561	-2.14%	MSE_KH	5.229	-3.20%	MSE_KH	3.46	-2.73%	MSE_KH	2.644	0.34%
MSE_PR	13.561	-2.14%	MSE_PR	5.229	-3.20%	MSE_PR	3.46	-2.73%	MSE_PR	2.644	0.34%

NOTE: I $= 0.3N(14, 25) + 0.7N(-6, 81)$; $\sigma_e^2 = 96$

Table A.7: Simulation results Under mixture distribution

(33). g=10, m=10			(34). g=10, m=25			(35). g=10, m=50			(36). g=10, m=100		
MSE_T	9.344	ARB	MSE_T	7.129	ARB	MSE_T	14.184	ARB	MSE_T	6.538	ARB
MSE_LW	9.995	6.97%	MSE_LW	7.415	4.01%	MSE_LW	14.618	3.06%	MSE_LW	6.489	-0.75%
MSE_KH	8.982	-3.87%	MSE_KH	7.113	-0.22%	MSE_KH	14.328	1.02%	MSE_KH	6.424	-1.74%
MSE_PR	8.985	-3.84%	MSE_PR	7.113	-0.22%	MSE_PR	14.328	1.02%	MSE_PR	6.424	-1.74%
(37). g=25, m=10			(38). g=25 m=25			(39). g=25, m=50			(40). g=25, m=100		
MSE_T	6.498	ARB	MSE_T	3.786	ARB	MSE_T	4.17	ARB	MSE_T	6.11	ARB
MSE_LW	6.933	6.69%	MSE_LW	3.917	3.46%	MSE_LW	4.154	-0.38%	MSE_LW	6.252	2.32%
MSE_KH	6.265	-3.59%	MSE_KH	3.761	-0.66%	MSE_KH	4.071	-2.37%	MSE_KH	6.19	1.31%
MSE_PR	6.266	-3.57%	MSE_PR	3.761	-0.66%	MSE_PR	4.071	-2.37%	MSE_PR	6.19	1.31%
(41). g=50, m=10			(42). g=50, m=25			(43). g=50, m=50			(44). g=50, m=100		
MSE_T	5.711	ARB	MSE_T	2.908	ARB	MSE_T	2.263	ARB	MSE_T	4.639	ARB
MSE_LW	6.031	5.60%	MSE_LW	3.074	5.71%	MSE_LW	2.321	2.56%	MSE_LW	4.681	0.91%
MSE_KH	5.465	-4.31%	MSE_KH	2.952	1.51%	MSE_KH	2.275	0.53%	MSE_KH	4.634	-0.11%
MSE_PR	5.465	-4.31%	MSE_PR	2.952	1.51%	MSE_PR	2.275	0.53%	MSE_PR	4.634	-0.11%
(45). g=100, m=10			(46). g=100, m=25			(47). g=100, m=50			(48). g=100, m=100		
MSE_T	5.275	ARB	MSE_T	2.77	ARB	MSE_T	1.486	ARB	MSE_T	3.06	ARB
MSE_LW	5.666	7.41%	MSE_LW	2.931	5.81%	MSE_LW	1.478	-0.54%	MSE_LW	3.095	1.14%
MSE_KH	5.143	-2.50%	MSE_KH	2.817	1.70%	MSE_KH	1.448	-2.56%	MSE_KH	3.064	0.13%
MSE_PR	5.143	-2.50%	MSE_PR	2.817	1.70%	MSE_PR	1.448	-2.56%	MSE_PR	3.064	0.13%

NOTE: II $= 0.3N(14, 9) + 0.7N(-6, 36)$; $\sigma_e^2 = 49$

Table A.8: Simulation results Under mixture distribution

(49). g=10, m=10			(50). g=10, m=25			(51). g=10, m=50			(52). g=10, m=100		
MSE_T	20.463	ARB	MSE_T	12.461	ARB	MSE_T	7.671	ARB	MSE_T	5.988	ARB
MSE_LW	23.689	15.77%	MSE_LW	12.711	2.01%	MSE_LW	7.809	1.80%	MSE_LW	5.96	-0.47%
MSE_KH	19.445	-4.97%	MSE_KH	11.711	-6.02%	MSE_KH	7.499	-2.24%	MSE_KH	5.841	-2.45%
MSE_PR	19.45	-4.95%	MSE_PR	11.712	-6.01%	MSE_PR	7.499	-2.24%	MSE_PR	5.841	-2.45%
(53). g=25, m=10			(54). g=25 m=25			(55). g=25, m=50			(56). g=25, m=100		
MSE_T	9.848	ARB	MSE_T	5.825	ARB	MSE_T	4.748	ARB	MSE_T	2.936	ARB
MSE_LW	12.058	22.44%	MSE_LW	6.306	8.26%	MSE_LW	4.893	3.05%	MSE_LW	2.911	-0.85%
MSE_KH	9.972	1.26%	MSE_KH	5.826	0.02%	MSE_KH	4.703	-0.95%	MSE_KH	2.853	-2.83%
MSE_PR	9.974	1.28%	MSE_PR	5.826	0.02%	MSE_PR	4.703	-0.95%	MSE_PR	2.853	-2.83%
(57). g=50, m=10			(58). g=50, m=25			(59). g=50, m=50			(60). g=50, m=100		
MSE_T	8.732	ARB	MSE_T	5.298	ARB	MSE_T	3.047	ARB	MSE_T	2.833	ARB
MSE_LW	10.62	21.62%	MSE_LW	5.781	9.12%	MSE_LW	3.15	3.38%	MSE_LW	2.871	1.34%
MSE_KH	8.82	1.01%	MSE_KH	5.35	0.98%	MSE_KH	3.028	-0.62%	MSE_KH	2.815	-0.64%
MSE_PR	8.821	1.02%	MSE_PR	5.35	0.98%	MSE_PR	3.028	-0.62%	MSE_PR	2.815	-0.64%
(61). g=100, m=10			(62). g=100, m=25			(63). g=100, m=50			(64). g=100, m=100		
MSE_T	8.694	ARB	MSE_T	4.457	ARB	MSE_T	2.788	ARB	MSE_T	1.634	ARB
MSE_LW	10.151	16.76%	MSE_LW	4.77	7.02%	MSE_LW	2.848	2.15%	MSE_LW	1.669	2.14%
MSE_KH	8.462	-2.67%	MSE_KH	4.418	-0.88%	MSE_KH	2.74	-1.72%	MSE_KH	1.637	0.18%
MSE_PR	8.462	-2.67%	MSE_PR	4.418	-0.88%	MSE_PR	2.74	-1.72%	MSE_PR	1.637	0.18%

NOTE: II $= 0.3N(14, 9) + 0.7N(-6, 36)$; $\sigma_e^2 = 96$

Table A.9: Simulation results of the unbalanced model with normal distribution

$\hat{\alpha}_1$			$\hat{\alpha}_6$			$\hat{\alpha}_{11}$			$\hat{\alpha}_{16}$			$\hat{\alpha}_{21}$		
MSE_T	10.069	ARB	MSE_T	13.497	ARB	MSE_T	6.347	ARB	MSE_T	5.782	ARB	MSE_T	4.569	ARB
MSE_LW	10.863	7.89%	MSE_LW	13.567	0.52%	MSE_LW	6.48	2.1%	MSE_LW	5.667	- 2.01%	MSE_LW	4.54	- 0.62%
MSE_KH	10.109	-0.4%	MSE_KH	13.02	-3.53%	MSE_KH	6.369	0.34%	MSE_KH	5.638	-2.49%	MSE_KH	4.525	- 0.96%
MSE_PR	10.109	-0.4%	MSE_PR	13.02	-3.53%	MSE_PR	6.369	0.34%	MSE_PR	5.638	-2.49%	MSE_PR	4.525	- 0.96%

Simulation setting: g=25,m=(10,10,10,10,10,15,15,15,15,15,25,25,25,25,25,35,35,35,35,35,40,40,40,40,40), $\sigma^2_\alpha = 81$, $\sigma^2_e = 49$.

Table A.10: Simulation results of the unbalanced model with mixture distribution

$\hat{\alpha}_1$			$\hat{\alpha}_6$			$\hat{\alpha}_{11}$			$\hat{\alpha}_{16}$			$\hat{\alpha}_{21}$		
MSE_T	10.521	ARB	MSE_T	6.76	ARB	MSE_T	6.658	ARB	MSE_T	5.207	ARB	MSE_T	3.373	ARB
MSE_LW	11.125	5.74%	MSE_LW	6.539	-3.26%	MSE_LW	6.898	3.6%	MSE_LW	4.923	-5.45%	MSE_LW	3.29	-2.45%
MSE_KH	10.57	0.47%	MSE_KH	6.314	- 6.59%	MSE_KH	6.858	2.99%	MSE_KH	4.891	-6.07%	MSE_KH	3.287	-2.55%
MSE_PR	10.57	0.54%	MSE_PR	6.314	-6.59%	MSE_PR	6.858	2.99%	MSE_PR	4.891	-6.07%	MSE_PR	3.287	-2.55%

Simulation setting: g=25,m=(10,10,10,10,10,15,15,15,15,15,25,25,25,25,25,35,35,35,35,35,40,40,40,40,40), $\Pi = 0.3N(14,9) + 0.7N(-6,36)$, $\sigma^2_e = 49$.

Simulation code

```
#################################################################################################
#
# R code to simulate one setting (g, m, sigma_{alpha}^2, sigma_{epsilon}^2, distribution)
#
################################################################################################

# Clear the workspace
rm(list=ls())

# Load required libraries
library(lme4)
library(Matrix)

# Set options
options(max.print = 100000)

# Specify the setting:
# The number of the clusters g=10, 25, 50, 100
g=10
# The  cluster size m1=10, 25, 50, 100
m1=25
# The true variance components
# For normal distributions, sigma_{alpha}^2 =25, 81 and sigma_{epsilon}^2 =49, 96
thetatilde=c(81,49)
# For mixture distributions, sigma_{alpha}^2 =25, 81 and sigma_{epsilon}^2 =49, 96
mu.true= c(14, −6) #mixture component means (14,−6)
sigma2.true = c(9, 36) #mixture component standard deviations (25,81) and (9,36)

# Compute useful derived quantities
m=rep(m1,g) # each cluster size for balanced data
N=sum(m) # total sample size

# Specify the true value of the regression parameter
beta=c(5,7,3) # beta = (intercept, between slope, within slope)

# Define vectors to store cluster means and related quantities
ybar=vector(length=g) # g vector for y cluster means
xbar=vector(length=g) # g vector for x cluster means
xbarone=vector(length=N) # N vector for x cluster means
ux=matrix(vector(length=g*(2)),g,2) # g times 2 matrix for between cluster covariates

# Generate the covariates
```

```
x=round(abs(rnorm(N,mean=0,sd=4))) # generate N-vector x

# Compute the x cluster means
for (xi in 1:g){
s=(xi-1)*m1 +1
r=xi*m1
xbar[xi]=mean(x[s:r])
xbarone[s:r]=xbar[xi]
}

# Construct between cluster covariate matrix ux with columns 1 and xbar
ux=cbind(1,xbar)
uxt=t(ux)  #transpose of matrix ux

# Compute matrix D and its inverse
DDDD= uxt %*% ux
DDDDinv=solve(DDDD)

# Construct the matrix X
X=matrix(vector(length=N*3),N,3)# Define the N times 3 matrix X
X[,1]=1
X[,2]=xbarone
X[,3]=x-xbarone

# Compute c1 and c2 for computation of H
c1=sum(xbar)/g
c2=sum(xbar^2)/g

# Construct the matrix Z
I_g=diag(rep(1,g)) # g times g identity matrix
onem=rep(1,m1) # m1 vector of ones
Z=kronecker(I_g,onem) # generate matrix Z

# Define vectors to store results
rep=3000 # Number of simulation replications
MSElwstorage=vector(length=rep)
MSEkhstorage=vector(length=rep)
MSEprstorage=vector(length=rep)
MSEprnstorage=vector(length=rep)
MSEtstorage=vector(length=rep)
M2KHstorage=vector(length=rep)
M2PRstorage=vector(length=rep)

##############################################################################################
# Simulation Loop
##############################################################################################

for(sr in 1:rep){
# Generate g random effects alpha from normal distribution with mean 0 and variance sigma_alpha
alpha=rnorm(g,mean=0,sd=sqrt(thetatilde[1]))

# Generate g random effects alpha from mixture distribution with mean mu.true
# and variance sigma2.true
#prob=rbinom(g, 1, 0.7)
#alpha=rnorm(g, mean=mu.true[prob+1], sd=sqrt(sigma2.true[prob+1]))

# Creating N vector of random effects
randomeffect=Z%*%alpha

# Generate true error e with mean 0 and variance sigma_e
e=rnorm(N,mean=0,sd=sqrt(thetatilde[2]))

# Generate response y
y=X%*%beta+randomeffect+e

# Store data in a data frame
```

```
mydata=data.frame(y=y,x1=X[,1],x2=X[,2],x3=X[,3],randomeffect=randomeffect,
cluster=rep(1:g,each=m1))

# Use the lmer4 package to fit the model
myanalysismodel=lmer(y~x1+x2+x3-1+(1|cluster),REML=TRUE,data=mydata)
aa=as.data.frame(VarCorr(myanalysismodel)) #extract results
sigma.alpha=sqrt(aa[1,4]) # obtain estimated sigma_alpha
sigma.e=sqrt(aa[2,4]) # obtain estimated sigma.e
betahat=fixef(myanalysismodel) # obtain estimated beta
# Extract estimates of between cluster regression parameters
zetabeta=betahat[1:2]

# Compute gamma and lambda
gamma=(m1*sigma.alpha^2)/(sigma.e^2+m1*sigma.alpha^2)
lambda=(sigma.e^2+m1*sigma.alpha^2)

# Computing y cluster means
for (xi in 1:g){
s=(xi-1)*m1 +1
r=xi*m1
ybar[xi]=mean(y[s:r])
}

# Compute EBLUP
alphahat1=gamma*(ybar[1]- ux[1,]%*%zetabeta)

# Compute different MSE applications
# Compute M1
M1_LW=(m1*sigma.alpha^2)/lambda*(sigma.e^2/m1+sigma.alpha^2*ux[1,]%*%DDDDinv%*%uxt[,1])
# Compute M2_PR
M2_PR=(2*sigma.e^2)/((m1-1)*lambda)
# Compute M2_KH
M2_KH=M2_PR*(1-ux[1,]%*%DDDDinv%*%uxt[,1])
# Compute MSE_KH
MSE_KH=M1_LW+M2_KH
# Compute MSE_PR
MSE_PR=M1_LW+M2_PR*2
# Compute MSE_LW
MSE_LW=(1/g)*(sigma.alpha^2/(c2-c1^2))*(c2+xbar[1]^2-2*xbar[1]*c1)+sigma.e^2/m1
# Compute MSE_T
MSE_t=(alphahat1-alpha[1])^2

# Store results
MSEkhstorage[sr]=MSE_KH
MSElwstorage[sr]=MSE_LW
MSEprstorage[sr]=MSE_PR
MSEtstorage[sr]=MSE_t
M2KHstorage[sr]=M2_KH
M2PRstorage[sr]=M2_PR
}

# Complete calculations of average MSE estimates
mean(MSEtstorage) #average MSE_T
mean(MSElwstorage)#average MSE_LW
mean(MSEkhstorage)#average MSE_KH
mean(MSEprstorage)#average MSE_PR
```

Bibliography

Anderson, T., 1969: *Statistical inference for Covariance Matrices with Linear Structure*. New York: Academic Press, 55–66.

Arora, V. and P. Lahiri, 1997: On the superiority of the bayesian method over the blup in small area estimation problems. *Statistica Sinica* 1053–1063.

Battese, G. E., R. M. Harter, and W. A. Fuller, 1988: An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association* **83**, 28–36.

Bickel, P. J., 1975: One-step Huber estimates in the linear model. *Journal of the American Statistical Association* **70**, 428–434.

Billingsley, P., 1968: *Convergence of Probability Measures*. New York: Wiley.

Billingsley, P., 2008: *Probability and measure*. John Wiley & Sons.

Casella, G. and R. L. Berger, 2002: *Statistical inference*, volume 2. Duxbury Pacific Grove, CA.

Cramér, H., 1999: *Mathematical Methods of Statistics*. Princeton: Princeton University Press.

Das, K., 1979: Asymptotic optimality of restricted maximum likelihood estimates for the mixed model. *Calcutta Statistical Association Bulletin* **28**, 125–142.

Edwards, A. W. F., 1984: *Likelihood*. CUP Archive.

Field, C. A., P. Zhen, and A. H. Welsh, 2008: Bootstrapping data with multiple levels of variation. *Canadian Journal of Statistics* **36**, 521–539.

Hartley, H. O. and J. N. Rao, 1967: Maximum-likelihood estimation for the mixed analysis of variance model. *Biometrika* **54**, 93–108.

Harville, D., 1991: Comment on robinson: estimation of random effects. *Statist. Sci* **6**, 42–44.

Harville, D. A., 1974: Bayesian inference for variance components using only error contrasts. *Biometrika* **61**(2), 383–385.

Harville, D. A., 1977: Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* **72**, 320–338.

Haslett, S. J., 2016: Small area estimation using both survey and census unit record data. *Analysis of Poverty Data by Small Area Estimation* 325–348.

Henderson, C. R., 1953: Estimation of variance and covariance components. *Biometrics* **9**, 226–252.

Henderson, C. R., 1975: Best linear unbiased estimation and prediction under a selection model. *Biometrics* **31**, 423–447.

Henderson, C. R., 1984: *Application of Linear Models in Animal Breeding*. Canada: University of Guelph Press.

Henderson, C. R., O. Kempthorne, S. R. Searle, and C. Von Krosigk, 1959: The estimation of environmental and genetic trends from records subject to culling. *Biometrics* **15**, 192–218.

Jiang, J., 1996: REML estimation: asymptotic behavior and related topics. *The Annals of Statistics* **24**, 255–286.

Jiang, J., 1997: Wald consistency and the method of sieves in REML estimation. *The Annals of Statistics* 1781–1803.

Jiang, J., 1998: Asymptotic properties of the empirical BLUP and BLUE in mixed linear models. *Statistica Sinica* **8**, 861–885.

Jiang, J., 1999: On unbiasedness of the empirical blue and blup. *Statistics & Probability Letters* **41**, 19–24.

Jiang, J., 2017: *Asymptotic Analysis of Mixed Effects Models: Theory, Applications, and Open Problems*. London: CRC Press.

Jiang, J., P. Lahiri, and S.-M. Wan, 2002: A unified jackknife theory for empirical best prediction with M-estimation. *The Annals of Statistics* **30**, 1782–1810.

Kackar, R. N. and D. A. Harville, 1981: Unbiasedness of two-stage estimation and prediction procedures for mixed linear models. *Communications in Statistics-theory and Methods* **10**, 1249–1261.

Kackar, R. N. and D. A. Harville, 1984: Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *Journal of the American Statistical Association* **79**, 853–862.

Lange, N. and L. Ryan, 1989: Assessing normality in random effects models. *The Annals of Statistics* **17**, 624–642.

Miller, J. J., 1977: Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. *The Annals of Statistics* **5**, 746–762.

Nobre, J. and J. Singer, 2007: Residual analysis for linear mixed models. *Biometrical Journal* **49**, 863–875.

Ortega, J. M. and W. C. Rheinboldt, 1973: *Iterative Solution of Nonlinear Equations in Several Variables*. New York: Academic Press.

Patefield, W., 1977: On the maximized likelihood function. *Sankhyā: The Indian Journal of Statistics, Series B* **39**, 92–96.

Patterson, H. D. and R. Thompson, 1971: Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**, 545–554.

Patterson, H. D. and R. Thompson, 1974: Maximum likelihood estimation of components of variance. *Proceeding Eight International Biometric Conference,Biometric Society* 197–207.

Pinheiro, J. and D. Bates, 2000: *Mixed-Effects Models in S and S-Plus*. New York: Springer Verlag.

Prasad, N. and J. Rao, 1990: The estimation of the mean squared error of small-area estimators. *Journal of the American Statistical Association* **85**, 163–171.

Rao, J. N., 2015: *Small area estimation*, volume 331. John Wiley & Sons.

Raudenbush, S. W. and A. S. Bryk, 2002: *Hierarchical linear models: Applications and data analysis methods*, volume 1. sage.

Richardson, A. and A. H. Welsh, 1994: Asymptotic properties of restricted maximum likelihood (REML) estimates for hierarchical mixed linear models. *Australian Journal of statistics* **36**, 31–43.

Robinson, G. K., 1991: That BLUP is a good thing: the estimation of random effects. *Statistical Science* **6**, 15–32.

Schützenmeister, A., U. Jensen, and H. Piepho, 2012: Checking normality and homoscedasticity in the general linear model using diagnostic plots. *Communications in Statistics - Simulation and Computation* **41**, 141–154.

Searle, S. R., 1987: *Linear Models for Unbalanced Data*. John Wiley & Sons.

Searle, S. R., G. Casella, and C. E. McCulloch, 1992: *Variance components*. New Jersey: John Wiley & Sons.

Thompson, W. A., 1962: The problem of negative estimates of variance components. *The Annals of Mathematical Statistics* **33**, 273–289.

Tippett, L. H. C., 1952: The methods of statistics. *The Methods of Statistics.* **2**, 222.

Verbeke, G. and G. Molenberghs, 2000: *Linear Mixed Models for Longitudinal Data*. New York: Springer Verlag.

Verbyla, A. P., 1990: A conditional derivation of residual maximum likelihood. *Australian Journal of Statistics* **32**, 227–230.

Xie, M. and Y. Yang, 2003: Asymptotics for generalized estimating equations with large cluster sizes. *The Annals of Statistics* **31**, 310–347.

Yoon, H.-J. and A. H. Welsh, 2020: On the effect of ignoring correlation in the covariates when fitting linear mixed models. *Journal of Statistical Planning and Inference* **204**, 18–34.